Information Processing and Energy Dissipation in Neurons

Lane McIntosh

Mathematics

University of Hawai'i at Manoa

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

April 2012

Thesis Committee

Susanne Still
Department of Information and Computer Science

George Wilkens
Department of Mathematics

Abstract

We investigate the relationship between thermodynamic and information theoretic inefficiencies in an individual neuron model, the adaptive exponential integrate-and-fire neuron. Recent work has revealed that minimization of energy dissipation is tightly related to optimal information processing, in the sense that a system has to compute a maximally predictive model. In this thesis we justify the extension of these results to the neuron and quantify the neuron's thermodynamic and information processing inefficiencies.

Contents

List of Figures

Acknowledgements

This project was first conceived by my advisor Susanne Still and Giacomo Indiveri at ETH Zurich; I am especially indebted to Professor Still for helping me at all stages of this research. I am also thankful to George Wilkens for all of the hours we spent pouring over stochastic differential equations, and his patience with questions that always seem trivial in retrospect. I also thank Dr. Robert Shaw, Jon Brown, John Marriott, and Eric Reckwerdt for their conversations and advice relating to this thesis, and Paul Nguyen for his explanation of "that $\cup$ symbol" on the first day of graduate real analysis.

I am also grateful to the National Science Foundation and the Mathematics Department at the University of Hawaii, Manoa for their financial support.

1

Introduction

1.1 Physical Systems and their Computations

It might be counterintuitive to think of arbitrary physical systems as computing models, but in many ways they do.(1, 2, 3, 4) Since physical systems change in response to forces from the external environment, we can consider the state of the physical system at any given time as an implicit model of what has happened previously.(4)

Some of these models are good and some are bad - for instance, measuring the displacement of a mass on a critically damped spring won't tell us much about how far we perturbed it hours ago.(5) On the other hand, we might suppose that neurons are very good at modeling, since they are able to communicate a finely detailed representation of our world solely through the generation of electrical action potentials.(6)

Broadly speaking, neuroscience seeks to understand how thought and behavior emerge from one particular physical system, the brain. Our brains have an incredible capacity for gathering sensory information and constructing behaviorally-relevant representations of the world via complex internal states. What exactly then are the structures of the brain whose states process information, and even more importantly, within this system, what are the physics of information processing?

In this thesis, we take the neuron to be the fundamental unit of information processing in the brain, and apply a new theoretical result that the inefficiencies in a system's abil-

ity to process information are exactly equivalent to the system's energetic inefficiencies. (4) One interpretation of this is that every system employs predictive inference insofar as it operates efficiently.(4)

This result bears a powerful implication for neuroscience - that features of neurons' energy consumption are dictated by their ability to represent information efficiently, and vice-versa. Since a significant fraction of energy consumption in the neuron orchestrates the propagation of action potentials,(7) we expect that characteristic signatures in neuron spike trains, like spike frequency adaptation for instance, might arise from the minimization of these information processing inefficiencies, or equivalently, the minimization of energy dissipation.

While the theoretical relationship that energy dissipation is equivalent to nonpredictive information has been proven true, its extension to the neuron requires care. For one, the equality only holds for Markov systems with well defined thermodynamic equilibria.(4) Towards this end, we must determine what the "state" of the neuron is exactly. Another consideration is that, although neurons are certainly energetically efficient, neurons are known to perform discrimination (8) and incidence timing (9) tasks that might differ significantly from predictive inference - perhaps performing these tasks well is more important to the organism than a strict minimization of energetic inefficiencies. An additional challenge is the bewildering diversity of neurons and the differing types of synaptic currents they are subjected to, which differ according to the function of the neuron in its particular neural circuit.(10)

In the following pages we will investigate whether or not neurons do in fact perform predictive inference by minimizing energetic and information processing inefficiencies, with these potential pitfalls in mind.

1.2 Plan of Action

This research bring together research on statistical mechanics, information theory, neuroscience, and neuromorphic engineering, and our background chapter will be appropriately broad. We will first discuss neurons and how they are typically dealt with

mathematically, before moving into a brief primer on probability theory. We will then introduce relevant concepts and theorems in information theory and a subset of statistical mechanics, far-from-equilibrium thermodynamics, which we will be using later. We then discuss historical results that have made connections between information theory and statistical mechanics. This will set the stage for an in-depth discussion of the theoretical results from (4), which form a basis for the applications in this paper. Each section in the background chapter solely contains past findings from other authors, even though at times we may take the stylistic liberty of discussing a result as if we were deriving it for the first time.

In the next chapter we will cover our methods, starting with a description of the neuron model we use in the paper, its relevance to actual neurons, the parameters typically used in the neuroscience literature, how the model was derived, and how we simulate it. It is here where we also make formal decisions as to the neuron's state and the protocol that drives it far-from-equilibrium. We also discuss in this chapter our methods for numerically solving the system using the Runge-Kutta method.

Lastly, we present our findings and discuss future experimental and analytical work.

2

Background

2.1 Biology and the Neuron

Neurons are excitable cells found in the brain and elsewhere in the nervous system from the mechanoreceptors in the bottoms of your feet to the interneurons of your brain's cortex, and have the extraordinary ability to capture and transmit information via stereotyped electrical impulses called action potentials or spikes.(10) While there is significant variation from neuron to neuron in spike timing, action potential shape, distribution of ion channel types that generate the spikes, and neurotransmitters that modulate and relay information between neurons, all of the information a neuron receives and distributes can be represented solely through it's digitized spike time series.(11)

But before we can begin our mathematical analysis of the task at hand, we must briefly familiarize ourselves with the biology of the neuron. Unlike canonical physical systems like a harmonic oscillator or an ideal gas compressed by a piston, a biological neuron has no *a priori* state, and we must use this qualitative knowledge to make a reasonable guess of what a neuron's "state" would be. Of course, while the complexity of a real neuron cannot be completely determined by just the voltage across the membrane, perhaps from an information theoretic perspective this is not an unreasonable assumption. This intuition about the simplified state of the neuron will also be integral to our choice of neuron model.

Furthermore, it is important for the biology of the neuron to constrain our neuron model, since otherwise we would have no idea as to whether or not our conclusions are a good approximation for what we would find *in vivo*. Before delving into a brief overview of how neurons are typically modeled both in theory and *in silica*, we also briefly review the concept of spike frequency adaptation, and how adaptation might be relevant to our application of (4).

2.1.1 The Brain

In an average 3 pound adult human brain, there are approximately 10^{11} neurons, each with roughly 7,000 connections to other neurons.(12, 13) In comparison to other sciences, knowledge about the brain has been painstakingly slow; although the brain has been regarded as the seat of mental activity since the second century, it was not even understood that the brain was comprised of cells until the nineteenth century.(14)

One particularly elusive piece of this puzzle had been identifying which parts of the brain give rise to mental activity, and how these structures integrate and process sensory information, perform inferences, generate cohesive thoughts, and lead to behavior. To this day, there is still considerable controversy over what the fundamental unit of computation is in the brain.(15, 16) Neurons interact with other neurons through excitatory and inhibitory synapses, which can significantly sculpt information as it passes from one neuron to another.(17) The vast majority of neurons receive input from many neighboring neurons, but ambiguity surrounds exactly where the integration of all this information takes place. Oftentimes the connections between neurons form a functional group that acts in synchrony; these groups, or circuits, are another candidate for the fundamental unit of computation in the brain.(15) Circuits in turn interact with other circuits to form large networks of neurons, which have also been argued to carry information not present at the individual neuron or circuit levels.(18)

Historically progress in neuroscience has been driven by the development of new technologies used to sample and image activity in the brain, and to this day most of these technologies have the capacity to only look at brain activity on small, disjoint subsets of spatial and temporal resolution, resulting in neuroscience communities that have very different views on the level of computation in the brain.

2. BACKGROUND

In this paper, we make the reasonable assumption that significant information is stored by action potentials, even at the single neuron level. However, we are still not free from controversy, since there is also disagreement as to how action potentials encode information; specifically, there is disagreement over whether these action potentials encode information via the firing rate $v(t)$ of the action potentials or via a precise temporal code $h(t)$ whereby the time between each action potential (called inter-spike intervals, or ISIs) is important.(19, 20, 21, 22, 23) Of course, choosing a single side is unnecessary, and there is evidence of neurons that use both strategies; for instance, photoreceptors in the retina are thought to be incidence detectors, and so temporal timing is of critical importance, while in area V1 of visual cortex there are neurons known to encode the orientation of edges in the visual field via spike rate.(6)

Action Potentials

Neurons, like all other cells throughout the body, are made distinct from the extracellular space by a lipid bilayer membrane that is impermeable to ions.(10) However, embedded in this membrane is the basis for all electrical signaling throughout the animal kingdom - ion channels. Ion channels are macromolecular pores that selectively transport ions back and forth through the cell membrane either passively (such that ions flow through the pore along the ion's concentration gradient) or actively (such that energy in the form of adenosine triphosphate, ATP, is expended to move ions against its concentration gradient), and are responsible for the production and transduction of all signals generated by and sent to the brain - from the contraction of muscles to the detection of sound waves.(24)

Action potentials, the substrate for theories of coding and computation in neurons, are rapid depolarizations of the cell membrane typically lasting < 1 ms generated by Na^+, K^+, and Cl^- ion channels.[1](24) The potential difference (or voltage) of the neuron's intracellular environment with respect to the extracellular space is given by the Goldman-Hodgkin-Katz equation as a function of the relative concentrations of these

[1]Note that in action potentials outside of the human brain, for instance in cardiac action potentials, Ca^{2+} ion channels are also involved, creating action potentials that are on the order of 100 times slower.

ions inside and outside the cell,

$$V_{\text{neuron}} = \frac{RT}{F} \ln \frac{P_{K^+}[K^+]_{\text{out}} + P_{Na^+}[Na^+]_{\text{out}} + P_{Cl^-}[Cl^-]_{\text{in}}}{P_{K^+}[K^+]_{\text{in}} + P_{Na^+}[Na^+]_{\text{in}} + P_{Cl^-}[Cl^-]_{\text{out}}}, \tag{2.1}$$

where [ion] is the concentration of the ion, P_{ion} is the permeability of the ion across the cell membrane, R is the ideal gas constant, T is the temperature, and F is Faraday's constant.(14) Note that a positive V then indicates that there are more positive ions outside of the cell than inside. At rest, a typical value of V_{neuron} would be around -70 mV, and in fact all excitable cells have a negative resting potential since there are Na^+-K^+ channels that actively pump positive sodium ions into the extracellular space and potassium ions into the cell (with 3 sodium ions leaving for every 2 potassium ions entering).(24) Since there are far more open potassium channels than open Na^+ or Cl^- channels, $P_{Na^+} >> \max\{P_{Na^+}, P_{Cl^-}\}$, and the small $[K^+]/[K^-]$ dominates V_{neuron}.(24)

An action potential occurs when the potential becomes depolarized sufficiently enough such that voltage-gated Na^+ ion channels open, letting positive sodium ions flow along their concentration gradient into the cell rapidly.(10) Since this increases the potential even more, even more voltage-gated sodium channels are opened; in this manner, an action potential is an all-or-nothing response. At the peak of the action potential, the sodium channels close and potassium channels open, letting the voltage fall back to resting potential.(10) After this occurs, there is a short "refractory" period during which the membrane must be recharged by the active ion channels, pumping Na^+ ions back into the extracellular space and K^+ ions back into the cell.(24)

Energy Consumption and Efficiency

With all of this shuttling of ions across the cell membranes of neurons, how substantial is the energy cost of transmitting information along a neuron? Although the human brain comprises only about 2% of our body mass, at rest the brain accounts for about 20% of oxygen consumption in the body and about 20% of the entire body's metabolism. (7, 25) Most of this disproportional consumption of energy comes from the Na^+-K^+ pump, which must break a phosphate bond of ATP for every $3Na^+/2K^+$ transported across a neuron's membrane.(26) During a single action potential event, this translates to 150×10^6 ATP molecules that are used up by the Na^+-K^+ pump alone, with a total

energetic cost of almost 400×10^6 ATP per action potential.(26)

Given the high energetic costs associated with information processing in neural tissue and evidence that evolution strongly minimizes energy consumption while maintaining the ability to adapt under changing environmental conditions, many theories of energy efficient neural coding have been developed in the last four decades.(27, 28, 29) Most of these codes seek efficiency by maximizing representational capacity[1] while reducing the average firing rate $v(t)$, minimizing redundancy, or using sparse and distributed codes.(30, 31)

Beyond governing how information is encoded, the need for energy efficiency in neural systems (without losing any signal to intrinsic noise) extends from the degree of inter-neuron wiring miniaturization in the brain to the distribution of ion channels in the cell membrane, and has been seen as the unifying principle of neural biophysics.(32)

Spike-frequency Adaptation

One mechanism by which neurons are thought to reduce $v(t)$ and operate more efficiently is spike-frequency adaptation.(33) Spike-frequency adaptation is the slow decrease in firing rate of a neuron after exposed to a steady stimulus current, and has been found in the neurons of a wide variety of organisms, from crustaceans to humans.(34) Adaptation in general is found in neurons and neural circuits on many different timescales for the purpose of increased dynamic range and sensitivity to small changes, and - at the loss of context - adaptation represents a maximization of the information the neuron or circuit transmits about its sensory inputs, increasing the efficiently of the neural code.(35, 36)

Spike-frequency adaptation in particular is a ubiquitous feature of spiking neurons that can be caused by a variety of mechanisms.(37) After an action potential occurs, there is a deep hyperpolarization (called an afterhyperpolarization, or AHP) during the action potential's refractory period; during tonic spiking, these AHPs can accumulate,

[1] Treating the action potential as a binary event, the representational capacity in bits per unit time of n neurons is $C(n, np) = \log_2 \left[\frac{n!}{(np)!(n-np)!} \right]$, where $p \in [0, 1]$ such that $np \in \mathbb{Z}^+$ is the number of neurons active.

slowing down the firing rate.(24) In addition to these AHP currents, it is also understood that currents generated by voltage-gated, high-threshold K^+ channels and the fast sodium current can also give rise to spike-frequency adaptation.(37, 38) Despite detailed biophysical knowledge of mechanisms underlying spike-frequency adaptation, the functional role of spike-frequency adaptation in computation is still relatively unknown.(34)

Given the possible relationship between spike-frequency adaptation and energy efficiency, we might want to investigate in the future if spike adaptation is perhaps an emergent phenomenon, stemming from the neuron's minimization of thermodynamic and information processing inefficiency.[1]

Neuron Models

Neuron models seek to accurately describe the electrical activity of the cell via a system of differential equations, and generally fall into one of two general categories - simple, mathematically tractable models of neuron spiking behavior and biophysically detailed models that simulate mechanisms underlying the neuron's activity.(39)

The history of theoretical models for neurons began with Louis Lapicque in 1907 with one of the simplest model neurons, the integrate-and-fire neuron,

$$I(t) = C_m \frac{dV_m}{dt}, \tag{2.2}$$

where $I(t)$ is the current, C_m is the membrane capacitance, and V_m is the potential difference across the membrane.(40) Since the lipid bilayer membrane is so thin, the accumulation of positive and negative charges on either side of the cell membrane leads to an electrical force that pulls oppositely-charged ions toward the other side, which can be described as a capacitance C_m.(14) *In vivo*, the movement of these ions across the membrane with associated charge Q creates a current according to

$$I(t) = \frac{dQ}{dt}. \tag{2.3}$$

The cell membrane however is also only semipermeable, and so has a membrane resistance R associated with it as ions are transported across the membrane. The ease at

[1] See **??**.

which the current crosses the membrane, or conductance g, is accordingly the inverse of this resistance, $g = 1/R$.(24)

On the opposite extreme of 2.2, detailed biophysical models take account of these conductances and ionic sources of current. In the earliest detailed biophysical model - the Hodgkin and Huxley model - the current $I(t)$ is broken up into component parts

$$I(t) = I_C(t) + \sum_k I_k(t), \tag{2.4}$$

where $I_C(t)$ is the portion of injected current that charges the capacitor (read: membrane) and the I_k are currents that pass through the sodium, potassium, and unspecified leak ion channels.(41) Each of these ion channels is associated with a conductance g_k, a resting potential E_k given by 2.1 with all permeabilities $P_j = 0$ for $j \neq k$[1], and gating variables m, n and h that determine the probability that a channel is open.(41) Looking back at the capacitative current $I_C(t)$ we can use the definition of capacitance, $C = Q/V$, and current, $I(t) = dQ/dt$, to find that

$$I_C(t) = C\frac{dV}{dt}. \tag{2.5}$$

Substituting our new expression for $I_C(t)$ into 2.2 and expanding the ion channel currents with the parameters described above, we find that the full Hodgkin and Huxley model is

$$\begin{aligned} C\frac{dV}{dt} &= I(t) - \left[g_{\mathrm{Na}} m^3 h(V - E_{\mathrm{Na}}) + g_{\mathrm{K}} n^4 (V - E_{\mathrm{K}}) + g_L(V - E_{\mathrm{L}})\right] && (2.6)\\ &= \frac{dm}{dt} = \alpha_m(V)(1-m) - \beta_m(V)m && (2.7)\\ &= \frac{dn}{dt} = \alpha_n(V)(1-n) - \beta_n(V)n && (2.8)\\ &= \frac{dh}{dt} = \alpha_h(V)(1-h) - \beta_h(V)h, && (2.9) \end{aligned}$$

where each α_i, β_i are empirical exponential functions of voltage.(42) Since the 1952 publication date of the Hodgkin-Huxley model, numerous additional models have been proposed that make compromises between these disparate categories of simple functional models and detailed biophysical ones (for a good overview see (41) or (43)). Of particular interest to us will be the adaptive exponential integrate-and-fire neuron

[1] This generalization of the Goldman-Hodgkin-Katz equation is known as the Nernst equation.

(see Chapter 3), an elaboration of the integrate-and-fire neuron 2.2 that demonstrates spike-frequency adaptation.(44)

In Silica Models

Given theoretical descriptions of the neuron's membrane and ion channels as resistors, capacitors, and batteries, it is no surprise that one could design an electrical circuit equivalent to the neuron in its electrical properties and implement this circuit *in silica*. Neuromorphic engineering seeks to design either analog or digital computer chips that emulate the behavior of real neurons, and to do this, engineers must consider the density of electrical components, the complexity and size of the circuit, the balance between analog and digital elements, and the energy efficiency and consumption of the circuit.(45)

Traditional computer implementations of neurons dissipate non-negligible amounts of energy and consume orders of magnitude more energy per instruction. In 1990, it was correctly estimated that computers use 10^7 times as much energy per instruction than the brain does.(46) This vast inefficiency led to the development of analog, low(er)-power silicon implementations of neurons, which still dissipate large amounts of power compared to the brain.(45) In addition to the drawback of needing to supply more power, implementations that dissipate large amounts of energy limit the density and miniaturization of its component parts on account of thermal noise.(45) In addition, neuromorphic prostheses intended to restore movement, hearing, or vision to patients face serious clinical challenges due to brain tissue damage caused by the dissipation of heat.(47)[1]

To overcome these issues, Giacomo Indiveri has recently developed *in silica* implementations of the adaptive exponential integrate-and-fire neuron that dramatically reduce the amount of dissipated energy, and so it will be possible in the future to experimentally verify the theoretical predictions of this paper.(48)

[1] In order to avoid damaging brain tissue, a 6×6 mm^2 array must dissipate less than 10mW of energy.(47)

2.2 Probability Theory

In this paper we will assume knowledge of basic probability theory, but we will mention a few key theorems and conventions that we will use later.

2.2.1 First Moment

We denote the average of X over a probability distribution p by the angle brackets $\langle X \rangle_p$. When the probability distribution is clear from the context, we will occasionally reference the average as $\langle X \rangle$.[1]

2.2.2 Normal Distribution

We say that a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ when X is normally distributed,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2.10}$$

where $p(x)$ is the probability density function.

2.2.3 Jensen's Inequality

Theorem (Jensen's Inequality) 2.2.1. *Assume that the function g is measurable and convex downward. Let the random variable x be such that $\langle |x| \rangle < \infty$. Then*

$$g(\langle x \rangle) \leq \langle g(x) \rangle. \tag{2.11}$$

2.3 Information Theory

Information as a mathematical quantity was first developed by Claude Shannon in his seminal work, "A Mathematical Theory of Communication," first published in 1948 (in fact, upon realizing the generality of the theory, later publications changed the article "a" to "the").(49, 50) Shannon represents an arbitrary (discrete) information source as a Markov process and then asks whether we can define a quantity that measures how much information the process produces, and at what rate. We think of information in this context as how interesting it is to discover the realization of the process. For instance, if a process $x(t) = 1$ with probability $\Pr(x = 1) = 1$ for all time t, then there

[1] Later this will especially occur when we average over the joint probability distribution of the process' state space $s(t)$ and the space of protocols $x(t)$.

is no information in discovering what x actually is at any time t. It is evident already that defining information will rely on knowledge of the probability distribution of the process, and not on the process itself per se.

2.3.1 Entropy

Suppose we have a discrete random variable that can take one of n states with probabilities $p_1, p_2, ..., p_n$. Then a measure of information H should satisfy

1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

This leads us to the important finding that

Theorem 2.3.1. *The only H satisfying the above three assumptions is of the form*

$$H = -K \sum_{i=1}^{n} p_i \log p_i, \tag{2.12}$$

where K is a positive constant.

Proof. Let H be a function of the probability distribution, $H(p_1, p_2, ..., p_n)$. From condition (2) we have $H\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) = f(n)$, where f is a monotonic increasing function of n. Applying condition (3), we can break down a choice from r^m equally likely possibilities into a series of m choices each from r equally likely possibilities, such that

$$f(r^m) = mf(r). \tag{2.13}$$

Similarly, we have $f(t^n) = nf(t)$. Furthermore, we can choose n arbitrarily large and find an m to satisfy

$$r^m \leq t^n < r^{m+1}. \tag{2.14}$$

Taking logarithms and dividing by $n \log r$, we then have

$$\frac{m}{n} \leq \frac{\log t}{\log r} \leq \frac{m}{n} + \frac{1}{n} \implies \left| \frac{m}{n} - \frac{\log t}{\log r} \right| < \epsilon, \tag{2.15}$$

where ϵ is arbitrarily small. Furthermore, since f is monotonic, $f(r^m) \leq f(t^n) \leq f(r^{m+1})$ implies that $mf(r) \leq nf(t) \leq (m+1)f(r)$, and so by dividing by $nf(r)$, we have

$$\frac{m}{n} \leq \frac{f(t)}{f(r)} \leq \frac{m}{n} + \frac{1}{n} \implies \left| \frac{m}{n} - \frac{f(t)}{f(r)} \right| < \epsilon. \tag{2.16}$$

Combining (2.15) with (2.16), we find that certainly

$$\left| \frac{f(t)}{f(r)} - \frac{\log t}{\log r} \right| \leq 2\epsilon. \tag{2.17}$$

Since ϵ is arbitrarily small, we are forced to have $f(t) = \frac{f(r)}{\log r} \log t$, where $\frac{f(r)}{\log r} > 0$ to satisfy the monotonicity of $f(t)$ required by condition (2). Since r is arbitrary, we let $\frac{f(r)}{\log r} = K$, where K is some positive constant.

Suppose we have $\sum_{i=1}^{n} n_i$ choices with equal probabilities $p_i = \frac{n_i}{\sum n_i}$, where $n_i \in \mathbb{Z}$. Then we have information measure

$$f\left(\sum n_i\right) = K \log \sum n_i. \tag{2.18}$$

Alternatively, we could break up the $\sum n_i$ choices into a choice from just n possibilities, with probabilities $p_1, ..., p_n$, and then, if the ith possibility was chosen, a choice from n_i with equal probabilities. This would then have information measure

$$H(p_1, ..., p_n) + K \sum p_i \log n_i. \tag{2.19}$$

However, by condition (3), both of these information measures must be equivalent,

$$K \log \sum n_i = H(p_1, ..., p_n) + K \sum p_i \log n_i, \tag{2.20}$$

and so using the properties of logarithms and the observation that $\sum p_i = 1$,

$$H = K \left[\sum p_i \log \sum n_i - \sum p_i \log n_i \right] \tag{2.21}$$

$$= -K \left[\sum p_i \log \left(\frac{n_i}{\sum n_i} \right) \right] = -K \sum p_i \log p_i. \tag{2.22}$$

□

We call this measure of information measure H entropy, and use $\log_2$ out of convention, measuring H in bits. Out of convenience, we take $K = 1$ and define the entropy of a discrete random variable X with probability distribution $p(x)$ as

$$H(X) = -\sum_{x \in X} p(x) \log p(x). \tag{2.23}$$

Since the "surprise value" or uncertainty of a process may change when another process becomes known, it is natural to consider conditional entropy $H(X|Y)$,

$$H(X|Y) := -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) \tag{2.24}$$

$$= -\langle \log p(x|y)\rangle_{p(x,y)}. \tag{2.25}$$

2.3.2 Relative Entropy

Relative entropy, also called Kullback-Leibler divergence, measures the distance[1] between two probability distributions $p(x)$ and $q(x)$,

$$D_{\mathrm{KL}}[p(x)||q(x)] = \left\langle \log \frac{p(x)}{q(x)} \right\rangle_{p(x)} \tag{2.26}$$

$$= \sum_x p(x)\log\frac{p(x)}{q(x)}. \tag{2.27}$$

Theorem (Information Inequality) 2.3.1. *Let X be a random variable with probability mass functions $p(x)$ and $q(x)$, where $x \in X$.(51) Then*

$$D_{KL}(p||q) \geq 0. \tag{2.28}$$

Proof. Let $A = \{x : p(x) > 0\}$ be the support of $p(x)$. Then

$$-D_{\mathrm{KL}}(p||q) = -\sum_{x\in A} p(x)\log\frac{p(x)}{q(x)} \tag{2.29}$$

$$= \sum_{x\in A} p(x)\log\frac{q(x)}{p(x)}. \tag{2.30}$$

By Jensen's inequality 2.2.3, we take the passage of the log under the summation such that

$$-D_{\mathrm{KL}}(p||q) \leq \log\sum_{x\in A} p(x)\frac{q(x)}{p(x)} \tag{2.31}$$

$$= \log\sum_{x\in A} q(x). \tag{2.32}$$

Next, since $A \subseteq X$ and the sum of any probability distribution over all of it's values must be 1, we must have

$$\log\sum_{x\in A} q(x) \leq \log\sum_{x\in X} q(x) = \log 1 = 0. \tag{2.33}$$

But since $-D_{\mathrm{KL}}(p||q) \leq 0$, we must have $D_{\mathrm{KL}}(p||q) \geq 0$. □

[1]Note however that the D_{KL} is not a *true* distance since it is not symmetric and does not satisfy the triangle inequality.

2.3.3 Mutual Information

Consider two random variables X and Y with probability mass functions $p(x)$ and $p(y)$, respectively, and joint probability mass function $p(x, y)$. Mutual information is then defined as the relative entropy between the joint distribution and the product of the two marginal distributions,(51)

$$\begin{aligned} I[X;Y] &:= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} && (2.34)\\ &= D_{\mathrm{KL}}[p(x,y)||p(x)p(y)] && (2.35)\\ &= \left\langle \log\frac{p(X,Y)}{p(X)p(Y)}\right\rangle_{p(X,Y)}. && (2.36)\end{aligned}$$

Since $p(x, y) = p(x)p(y)$ if X and Y are independent, mutual information can be interpreted as measuring what you can learn about X from Y, and vice-versa; if X and Y are independent, then there is zero mutual information, and we can learn nothing about X from Y.

Recalling that the joint distribution is related to conditional probability by $p(x|y)p(y) = p(x, y)$, we can reformulate $I[X;Y]$ in terms of entropies:

$$\begin{aligned} I[X;Y] &= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} && (2.37)\\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x) + \sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) && (2.38)\\ &= -\sum_{x\in X} p(x)\log p(x) + \sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) && (2.39)\\ &= H(X) - H(X|Y) && (2.40)\\ &= H(X) + H(Y) - H(X,Y), && (2.41)\end{aligned}$$

provided that the joint entropy $H(X, Y)$ is well-defined.(52) Note that mutual information is symmetric, so $I[X_1; X_2] = H(X) - H(X|Y) = H(Y) - H(Y|X) = I[X_2; X_1]$.

Corollary 2.3.1. *Let X and Y be two random variables. Then*

$$I[X;Y] \geq 0. \tag{2.42}$$

Proof. By our definition 2.34, we have

$$I[X;Y] = D_{\mathrm{KL}}[p(x,y)||p(x)p(y)], \tag{2.43}$$

which is greater or equal to zero by 2.3.2. □

2.3.4 Inequalities

In addition to 2.3.2 and 2.3.3, from (51, 52) we also have the following inequalities:

- $H(X) + H(Y) \geq H(X, Y)$,
- $H(X, Y) \geq H(X)$,
- $H(X) \geq H(X|Y)$, and
- $H(X) \geq 0$.

2.4 Far-from-Equilibrium Thermodynamics

In our paradigm, we consider a stochastic physical system with state vector $s(t)$ driven from equilibrium by some process $x(t)$ over a discrete time scale $t \in \{0, 1, \ldots, \tau\}$. Since the physical system is stochastic, its state s given the protocol x is described by the probability distribution $p(s|x)$, and we let the time evolution of the system be given by a discrete-time Markov process with transition probabilities $p(s_t|s_{t-1}, x_t)$. For a Markov process (53), the transition to state s_t depends only on the preceding state s_{t-1}, such that

$$p(s_t|s_{t-1}, x_t) = p(s_t|s_{t-1}, s_{t-2}, \ldots, s_0, x_t). \tag{2.44}$$

During this process $x(t)$, we perform work W on the system, which in turn absorbs heat δQ. Additionally, we let the physical system be embedded in an environment of temperature T; usually this amounts to coupling the system to a heat bath of constant temperature such that any heat gained by the system is immediately whisked away. While classical thermodynamics studies the relationships between these quantities of work and various forms of energy (most prominently, thermal and free energy) where $p(s|x)$ is an equilibrium distribution determined solely by x, far-from-equilibrium thermodynamics is the study of these relationships in systems driven from equilibrium where $p(s|x)$ explicitly depends on the dynamics and history of the system's trajectory through state space.(54)

Most results in far-from-equilibrium thermodynamics follow from specific statements of the second law of thermodynamics, which asserts that systems tend to equilibrium,

and various theorems from probability theory.(55) Although the second law of thermodynamics was first stated by Sadi Carnot in 1824, modern thermodynamics starts with Rudolf Clausius' restatement of the second law in 1855.(56) Clausius demonstrated that for a cyclic process,

$$\oint \frac{\delta Q}{T} \leq 0, \tag{2.45}$$

where δQ is the amount of heat absorbed by the system, T is the temperature of the system, and the inequality is strict in the case where the process is irreversible. Furthermore, Clausius provided the first definition of entropy S; letting S be a state function that satisfies $dS = \delta Q/T$, Clausius then used (2.45) to state that entropy changes obey

$$\Delta S \geq \int \frac{\delta Q}{T}. \tag{2.46}$$

Realizing the statistical nature of this tendency towards disorder or "mixedupness," as the system absorbs heat, Ludwig Boltzmann in the same century reformulated entropy S in terms of the probabilities p_i that a system has states s_i,

$$-k_B \sum_i p_i \log p_i, \tag{2.47}$$

where k_B is the Boltzmann constant.

Departing from this historical narrative, let us formally introduce the concepts of work W and free energy F. In the context of thermodynamics, work was first defined as the mechanical equivalent of heat ($dW \propto dQ$), which was then later refined to accommodate potential energy E,

$$\delta W = \delta Q - dE, \tag{2.48}$$

where intuitively we see that performing work on the system is equivalent to adding heat into the system while accounting for the system's change in potential energy. The concept of free energy F naturally arises from the desire to quantify the energy in a system that is available for performing work, and was first given by Hermann von Helmholtz in 1882 as

$$F = E - TS, \tag{2.49}$$

although looser conceptions of free energy date back to the idea of affinity in the thirteenth century, when chemists sought an explanation for the force that caused chemical

reactions.(57)

Using these definitions of work and free energy (and the linearity of integration), we can restate the second law of thermodynamics (2.46) as

$$\Delta S \geq \int \frac{dU - \delta W}{T} \implies \int \delta W \geq \Delta U - T\Delta S \implies W \geq \Delta F, \tag{2.50}$$

that is, the work we perform on the system is never less than the change in free energy between the equilibrium state we started with and the equilibrium state we ended with.(55)

One key caveat to (2.72) is that it applies only to *macroscopic* systems; when we move to the microscopic realm we must interpret $W \geq \Delta F$ statistically, such that

$$\langle W \rangle \geq \Delta F, \tag{2.51}$$

where we are averaging W over the distribution of possible protocol trajectories $\vec{x}$.(55) In addition to these statistical considerations, we must also pay special attention to how we define free energy in a system arbitrarily far from equilibrium. But first, what does it mean to be in equilibrium?

2.4.1 Equilibrium

Given an initial condition $x(0) = x_0$ of our protocol, the physical system starts in equilibrium if the probability distribution $p(s_0|x_0)$ of the initial state s_0 given x_0 is Boltzmann distributed according to

$$p_{\text{eq}}(s_0|x_0) = e^{-\beta[E(s_0,x_0)-F(x_0)]}, \qquad \beta = \frac{1}{k_B T}, \tag{2.52}$$

where $E(s_0, x_0)$ is the energy of state s_0 and protocol x_0, and the free energy $F(x_0)$ denotes the energy in the system available for performing work at equilibrium.(58)

2.4.2 Free Energy

Recall that at equilibrium we had free energy $F = E - TS$ where E was the internal energy $E(s, x)$ of the system, T is temperature, and S is entropy. Since the amount of energy available to perform work in a system far from equilibrium not only depends

on the protocol $x(t)$ but also on the probability distribution $p(s|x)$, we introduce the concept of non-equilibrium free energy

$$F_{\text{neq}}[p(s|x)] := k_B T \, D_{\text{KL}}[p(s|x)||p_{\text{eq}}(s|x)], \tag{2.53}$$

which we can interpret as the thermodynamic difference between the non-equilibrium distribution of states and the distribution of systems states at equilibrium.(59)

Following (60), we also introduce generalized free energy F_G, which we will define as the difference between the average system energy $E(s,x)$ and the temperature-scaled entropy of the system,

$$F_G = \langle E(s,x) \rangle_{p(s|x)} - TS, \tag{2.54}$$

where entropy $S = -k_B \sum p(s|x) \log p(s|x)$. We demonstrate below that this in fact is equal to the sum of the equilibrium and non-equilibrium free energies.(4)

Theorem 2.4.1. *Let the generalized free energy*

$$F_G[p(s|x)] = \langle E(s,x) \rangle_{p(s|x)} - TS. \tag{2.55}$$

Then $F_G[p(s|x)]$ is the sum of equilibrium free energy $F(x)$ and nonequilibrium free energy $F_{neq}[p(s|x)]$.

Proof. Applying the definition of expectation, the generalized free energy becomes

$$\begin{aligned} F_G[p(s|x)] &= \langle E(s,x) \rangle_{p(s|x)} - TS & (2.56) \\ &= \sum p(s,x) E(s,x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)}. & (2.57) \end{aligned}$$

Consider the term $0 = F(x) - F(x)$; noting that $\sum_s p(s|x) = 1$ and $F(x)$ is not a function of s, we equivalently have $0 = F(x) - \sum F(x) p(s|x)$. Then 2.56 becomes by linearity

$$\begin{aligned} F_G &= \sum p(s,x) E(s,x) + F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} - \sum F(x) p(s|x) & (2.58) \\ &= F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} + \sum p(s,x) \left[E(s,x) - F(x) \right] & (2.59) \\ &= F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} - k_B T \sum p(s,x) \left[-\frac{1}{k_B T} [E(s,x) - F(x)] \right] & (2.60) \\ &= F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} - k_B T \sum p(s,x) \log e^{-\frac{1}{k_B T}[E(s,x) - F(x)]}. & (2.61) \end{aligned}$$

But $p_{\text{eq}} = e^{-\frac{1}{k_B T}[E(s,x)-F(x)]}$, and so

$$\begin{aligned} F_G &= F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} - k_B T \sum p(s,x) \log p_{\text{eq}}(s|x) \quad (2.62) \\ &= F(x) + k_B T \langle \log p(s|x) \rangle_{p(s|x)} - k_B T \langle \log p_{\text{eq}}(s|x) \rangle_{p(s|x)}, \quad (2.63) \end{aligned}$$

which by the property of logarithms gives

$$F_G = F(x) + k_B T \left\langle \log \frac{p(s|x)}{p_{\text{eq}}} \right\rangle_{p(s|x)}, \quad (2.64)$$

where the right-most term is the relative entropy $D_{\text{KL}}[p(s|x)||p_{\text{eq}}(s|x)]$. Since $F_{\text{neq}}[p(s|x)]$ is exactly $k_B T\, D_{\text{KL}}[p(s|x)||p_{\text{eq}}(s|x)]$, we must then have

$$F_G = F(x) + F_{\text{neq}}[p(s|x)]. \quad (2.65)$$

□

2.4.3 Dissipation versus Excess Work

Using this new distinction between free energy and generalized free energy, (4) establishes two new quantities, excess and dissipated work. We follow (4) and define excess work[1] to be

$$W_{\text{ex}} = W - \Delta F, \quad (2.66)$$

where $\Delta F = F[x_\tau] - F[x_0]$ is the change in free energy during a protocol that lasts τ long. Note that $F[x_\tau]$ and $F[x_0]$ are in fact equilibrium quantities; ΔF would be the work done if the protocol $x(t)$ changed infinitely slowly such that $p(s_{t-1}|x_t) = p(s_t|x_t)$ and each $p(s_t|x_t)$ is an equilibrium distribution for all time $0 \leq t \leq \tau$.(4)

In contrast, dissipated work is the average work that is irretrievably lost,(4)

$$W_{\text{diss}} = W - \Delta F_G, \quad (2.67)$$

where $\Delta F_G = F_G[p_\tau(s|x_\tau)] - F_G[p_0(s|x_0)]$ is the change in generalized free energy.[2]

[1] There is some confusion in the literature regarding these definitions; several authors call our excess work their dissipated work.

[2] The notation $p_t(s|x_t)$ is interchangeable with $p(s_t|x_t)$. These are equivalent to the transition probabilities from state s_{t-1} averaged over all possible states s_{t-1}, $\langle p(s_\tau|s_{\tau-1}, x_t) \rangle_{p(s_{\tau-1}|x_\tau)}$.

If we look at the average dissipation $\langle W_{\text{diss}} \rangle$ over the change in protocol from x_0 to x_τ, we note that

$$\begin{aligned} \langle W_{\text{diss}} \rangle &= \langle W \rangle - \Delta F - F_{\text{neq}}[p_\tau] && (2.68) \\ &= \langle W_{\text{ex}} \rangle - F_{\text{neq}}[p_\tau] && (2.69) \\ &\leq \langle W_{\text{ex}} \rangle, && (2.70) \end{aligned}$$

where p_τ is the probability distribution $p(s|x_\tau)$.(4)

2.4.4 Detailed Balance

Detailed balance is essentially a statement of microscopic reversibility at equilibrium, such that for any two states s_a and s_b,

$$p_{\text{eq}}(s_a \to s_b) = p_{\text{eq}}(s_b \to s_a). \tag{2.71}$$

Let a system with equilibrium state s_0 be driven through some path to the state s_τ under the change of protocol from x_0 to x_τ. We follow (54) and define work

$$W = \sum_{t=0}^{\tau-1} E(s_t, x_{t+1}) - E(s_t, x_t) \tag{2.72}$$

and heat

$$Q = \sum_{t=1}^{\tau} E(s_t, x_t) - E(s_{t-1}, x_t). \tag{2.73}$$

Then $W + Q = E(s_\tau, x_\tau) - E(s_0, x_0) = \Delta E$. With this in mind, we will now apply detailed balance to the probabilities of the forward and reverse paths through state space, $P_F(\vec{s}|\vec{x})$ and $P_R(\vec{s}|\vec{x})$, respectively.

Assuming detailed balance, we first observe that

$$\frac{P_F(\vec{s}|\vec{x})}{P_R(\vec{s}|\vec{x})} = \frac{p_{\text{eq}}(s_0|x_0)}{p_{\text{eq}}(s_\tau|x_\tau)} \prod_t \frac{p(s_t|s_{t-1}, x_t)}{p(s_{t-1}|s_t, x_t)}, \tag{2.74}$$

to which we can apply the definitions of equilibrium 2.52 and ΔE, seeing that

$$\begin{aligned} \frac{P_F(\vec{s}|\vec{x})}{P_R(\vec{s}|\vec{x})} &= e^{\beta(\Delta E - \Delta F)} \prod_t \frac{p(s_t|s_{t-1}, x_t)}{p(s_{t-1}|s_t, x_t)} && (2.75) \\ &= e^{\beta(\Delta E - \Delta F)} \prod_t \frac{p_{\text{eq}}(s_t|x_t)}{p_{\text{eq}}(s_{t-1}|x_t)}, && (2.76) \end{aligned}$$

where $\Delta F = F[x_\tau] - F[x_0]$ as before. Taking the definition of the equilibrium distribution again, we obtain

$$\begin{aligned} \frac{P_F(\vec{s}|\vec{x})}{P_R(\vec{s}|\vec{x})} &= e^{\beta(\Delta E - \Delta F)} \cdot e^{-\beta \sum_t [E(s_t,x_t) - E(s_{t-1},x_t)]} && (2.77) \\ &= e^{\beta(Q+W-\Delta F)} \cdot e^{-\beta Q} && (2.78) \\ &= e^{\beta W - \Delta F} = e^{\beta W_{\text{ex}}}, && (2.79) \end{aligned}$$

that is, the degree of reversibility of the path through state space is equal to $e^{\beta W_{\text{ex}}}$.

2.4.5 Jarzynski's Work Relation

We can now apply the above result to prove the following work relation.(58)

Theorem (Jarzynski's Work Relation) 2.4.1. *Let W be work, ΔF be the free energy difference $F[x_\tau] - F[x_0]$, and $\beta = \frac{1}{k_B T}$.*[1] *Then*

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \tag{2.80}$$

Proof. The average $\langle e^{-\beta W} \rangle$ is taken over all possible paths through state space over all protocols from x_0 to x_τ; i.e., we are averaging over P_F. Then by applying 2.77 we have

$$\begin{aligned} \langle e^{-\beta W} \rangle_{P_F} &= \langle e^{-\beta W} \frac{P_F}{P_R} \rangle_{P_R} && (2.81) \\ &= \langle e^{-\beta W} e^{\beta W_{\text{ex}}} \rangle_{P_R} && (2.82) \\ &= \langle e^{-\beta (W - W_{\text{ex}})} \rangle_{P_R}. && (2.83) \end{aligned}$$

But $W_{\text{ex}} = W - \Delta F$, and so $\langle e^{-\beta W} \rangle_{P_F} = \langle e^{-\beta \Delta F} \rangle_{P_R}$. Furthermore, β is just a constant and ΔF only depends on equilibrium values $F[x_\tau]$ and $F[x_0]$, so the expectation brackets vanish, leaving $\langle e^{-\beta \Delta F} \rangle_{P_R} = e^{-\beta \Delta F}$. Hence $\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$.

□

This result nicely constrains the possible distributions of work values W even when the system is driven far from equilibrium; the theorem also implies that we can measure equilibrium free energy differences from the behavior of the system far from equilibrium.(55)

[1] Note that T is not the temperature during the process, which very well might be far from equilibrium where temperature is not defined. Rather, T is the temperature of the heat bath coupled to the system.

2.4.6 Crooks' Fluctuation Theorem

Crooks' fluctuation theorem states that

$$\frac{\rho_F(+W)}{\rho_R(-W)} = e^{\beta(W-\Delta F)}. \tag{2.84}$$

2.5 Bridging Information Theory and Statistical Mechanics

Historically there have been several results relating information theory and statistical mechanics, most notably by E.T. Jaynes in the 1950's and Rolf Landauer in the 1960's. As we will see in the next section, these results are extended by the recent findings of Still, Sivak, Bell, and Crooks.(4)

2.5.1 E.T. Jaynes

In 1957, E.T. Jaynes published two landmark papers on the subject of information theory and statistical mechanics, in which he reinterpreted statistical mechanics as a form of statistical inference rather than a physical theory, where the only physical aspect of statistical mechanics lies in the correct determination of a system's states.(61, 62) Partially driven by the inability of classical thermodynamics to generalize to nonequilibrium conditions, Jaynes' approach removed the need for additional assumptions like ergodicity, metric transitivity, and equal *a priori* probabilities.

Suppose we have a system with n discrete energy levels $E_i(\alpha_1, \alpha_2, \dots)$, where each α_i is an external parameter such as volume, gravitational potential, or position of optical laser trap. Then if we only know the average energy $\langle E \rangle$ of the system, we cannot solve for the probabilities p_i such that

$$\langle E(\alpha_1, \alpha_2, \dots) \rangle = \sum_{i=1}^{n} p_i E_i(\alpha_1, \alpha_2, \dots) \tag{2.85}$$

unless our knowledge is augmented by $(n-2)$ more conditions, not including the normalization condition that

$$\sum_{i=1}^{n} p_i = 1. \tag{2.86}$$

This problem of choosing the probabilities is inherently a statistical one,[1] and if we are to consider probabilities as a reflection of our ignorance, then a good choice of p_i is that which correctly represents our state of knowledge while remaining maximally unbiased or uncertain with respect to what we do not know. Since entropy $H(p_1, p_2, \dots, p_n) = -\sum p_i \log p_i$ is a unique, unambiguous criterion for this amount of uncertainty (see 2.3.1), we can infer the probabilities p_i by maximizing their entropy subject to what is known.

Subject to the constraints 2.85 and 2.86, we can then maximize entropy by introducing the Lagrangian function Λ with multipliers λ and μ such that

$$\Lambda(p_i, \lambda, \mu) = -\sum_{i=1}^{n} p_i \log p_i - \lambda \left(\sum_{i=1}^{n} p_i - 1 \right) - \mu \left(\sum_{i=1}^{n} p_i E_i(\alpha_1, \alpha_2, \dots) - \langle E(\alpha_1, \alpha_2, \dots) \rangle \right). \quad (2.87)$$

Setting $\nabla_{p_i,\lambda,\mu} \Lambda(p_i, \lambda, \mu) = 0$, we find that we must have

$$\frac{\partial}{\partial p_i} = -(\log p_i + 1) - \lambda - \mu E_i(\alpha_1, \alpha_2, \dots) = 0. \quad (2.88)$$

Letting $\lambda_1 = \lambda - 1$, 2.88 then gives us our choice of each p_i as

$$p_i = e^{-\lambda_1 - \mu E_i(\alpha_1, \alpha_2, \dots)}. \quad (2.89)$$

Substituting this choice in to our constraints 2.85 and 2.86, Jaynes' then derives the canonical equilibrium distribution

$$p_i = e^{-\beta E_i(\alpha_1, \alpha_2, \dots)}. \quad (2.90)$$

Proceeding with the usual definition 2.49 of free energy $F(T, \alpha_1, \alpha_2, \dots) = U - TS$, Jaynes proceeds to show that

$$F(T, \alpha_1, \alpha_2, \dots) = k_B T \log Z(T, \alpha_1, \alpha_2, \dots), \quad (2.91)$$

where the partition function $Z(T, \alpha_1, \alpha_2, \dots) = \sum e^{-\beta E_i}$, and so we must have thermodynamic entropy

$$S = \frac{\partial F}{\partial T} = -k_B \sum p_i \log p_i. \quad (2.92)$$

[1] In fact, this is a very old statistical problem, dating back to Pierre-Simon Laplace's "Principle of Insufficient Reason" in the early 1800's.

In addition to being equal mathematically aside from the Boltzmann constant k_B, the thermodynamic and information entropy terms are conceptually identical from this vantage point, since both are essentially statements of statistical inference.

2.5.2 Landauer's Principle

In some sense Rolf Landauer advocated the converse of Jaynes' thesis; while E.T Jaynes revealed statistical mechanics as a form of inference, Rolf Landauer emphasized the physical nature of information. In 1961, Landauer connected the erasure of information $\mathcal{I}_e = H[s_0|x_0] - H[s_\tau|x_\tau]$ with an energy cost of $k_B T \ln 2$ per bit[1].(63) Initially appearing in an article of the IBM journal, Landauer's intent was to identify the lower limit of energy consumption in computing machines; however, this finding was remarkable for suggesting that arbitrary physical systems carried out computations by means of transitions between their states.(64) Extending this in subsequent papers, Landauer boldly argued that information is always tied to a physical representation - whether that representation be the electrical state of a paired transistor and capacitor in a computer's memory cell, a DNA configuration, the up or down spin of an electron, or the state of a neuron - and is never purely abstract.(1)

Given the first law of thermodynamics, which states that the energy of a closed system is conserved,

$$dU = \delta Q - \delta W^2, \tag{2.93}$$

Landauer observed that the erasure of information $\mathcal{I}_e$ requires heat to flow out of the system, such that

$$-\beta\langle Q\rangle = \mathcal{I}_e + \beta\langle W_{\text{diss}}\rangle \geq \mathcal{I}_e. \tag{2.94}$$

2.6 Thermodynamics of Prediction

Keeping the conventions used in 2.4 we now review the results of Still et al. 2012, which draw explicit relationships between thermodynamic dissipation W_{diss} and information theoretic inefficiencies, uniting the fields of information theory and statistical

[1] In the 1950's von Neumann proposed that any logical operation costs at least $T \ln 2$. However, Landauer showed that when computation is done reversibly, no dissipation occurs, and in fact the only theoretical energy cost of computation lies in the erasure of information.

[2] Here δQ and δW are the infinitesimal amounts of heat and work done by the system, respectively.

mechanics.(4)

Consider a physical system in thermodynamic equilibrium with states $s_t \in \{s_0, \ldots, s_\tau\}$ that is driven through state space by a protocol $x_t \in \{x_0, \ldots, x_\tau\}$ governed by some probability distribution $P_X(x_0, \ldots x_\tau)$. Suppose as in 2.4 that the dynamics of the system states s_t are described by the discrete Markov transition probabilities $p(s_t|s_{t-1}, x_t)$, such that a change in the driving signal $x_0 \to x_1$ forces the system out of equilibrium from $s_0 \to s_1$ according to $p(s_1|s_0, x_1)$.[1] We also (as before) couple the system to a heat bath at constant temperature T so that it can dissipate any heat gain δQ.

As before, the equilibrium distribution $p_{\text{eq}}(s|x_t) := e^{-\beta(E(s,x_t)-F[x_t])}$ and the probability $p(s_t|x_t)$ of seeing a state s_t after the system adjusts to x_t is given by the average of transitions from all possible s_{t-1} to s_t, $\langle p(s_t|s_{t-1}, x_t)\rangle_{p(s_{t-1}|x_t)}$. We also note that the probability of a specific path S through state space, conditional on a protocol $X = \{x_0, \ldots x_\tau\}$, is

$$P_{S|X} = p_{\text{eq}}(s_0|x_0) \prod_{t=1}^{\tau} p(s_t|s_{t-1}, x_t), \tag{2.95}$$

and the joint probability of state and protocol paths S and X is

$$P_{S,X} = p(x_0) p_{\text{eq}}(s_0|x_0) \prod_{t=1}^{\tau} p(x_t|x_0, \ldots, x_{t-1}) p(s_t|s_{t-1}, x_t). \tag{2.96}$$

We now define two information theoretic terms. Let the system's instantaneous *memory* be the mutual information between the system's current state s_t and the protocol step x_t,

$$I_{\text{mem}}(t) = I[s_t; x_t] := \left\langle \log \left[\frac{p(s_t|x_t)}{p(s_t)} \right] \right\rangle_{p(s_t|x_t)p(x_t)}. \tag{2.97}$$

Since the dynamics of s_t are determined by x_t and s_{t-1}, presumably the system state contains some information about the protocol. The degree to which the system can then predict the next instantiation of the protocol is given by the instantaneous *predictive power*,

$$I_{\text{pred}}(t) = I[s_t; x_{t+1}] := \left\langle \log \left[\frac{p(s_t|x_{t+1})}{p(s_t)} \right] \right\rangle_{p(s_t|x_{t+1})p(x_{t+1})}. \tag{2.98}$$

[1] It is worth noting that the conditional distribution $p(s_{t-1}|x_t)$ of s_{t-1} immediately after $x_{t-1} \to x_t$ and the conditional distribution $p(s_t|x_t)$ describing the system after it adjusts to the signal change $x_{t-1} \to x_t$ are not the same in general, and neither is an equilibrium distribution.

With these defined, a natural measure of the system's inefficiency is the amount of information s_t retains about x_t that is not generalizable to x_{t+1}. We call this information processing inefficiency the systems's instantaneous *nonpredictive information*

$$I_{\text{nonpred}}(t) = I_{\text{mem}}(t) - I_{\text{pred}}(t). \tag{2.99}$$

Recalling how we quantified generalized thermodynamic inefficiency as W_{diss} in 2.67, we are now prepared to ask if there is a relationship between the system's thermodynamic inefficiencies and the system's information theoretic inefficiencies.

Theorem 2.6.1. *Given the definitions and setup above, the instantaneous nonpredictive information scaled by k_BT is exactly the thermodynamic inefficiency averaged over all possible paths through the state space of the system and over all protocols,*

$$I[s_t; x_t] - I[s_t; x_{t+1}] = \beta\langle W_{diss}(x_t \to x_{t+1})\rangle. \tag{2.100}$$

Proof. Recalling our formulation of mutual information in terms of conditional entropy 2.37, we have

$$\begin{aligned} I[s_t; x_t] - I[s_t; x_{t+1}] &= H(s_t) - H(s_t|x_t) - [H(s_t) - H(s_t|x_{t+1})] && (2.101)\\ &= H(s_t|x_{t+1}) - H(s_t|x_t). && (2.102)\end{aligned}$$

Next, we can use our definition of generalized free energy 2.54 to rewrite the entropies above as thermodynamic quantities. Since generalized free energy

$$\begin{aligned} F_G[p(s|x)] &= \langle E(s,x)\rangle_{p(s|x)} - TS && (2.103)\\ &= \langle E(s,x)\rangle_{p(s|x)} + k_BT\sum p(s|x)\log p(s|x) && (2.104)\\ \beta F_G[p(s|x)] &= \beta\langle E(s,x)\rangle_{p(s|x)} + \sum p(s|x)\log p(s|x), && (2.105)\end{aligned}$$

we can average over $p(x)$ to obtain

$$\begin{aligned} \beta\langle F_G[p(s|x)]\rangle_{p(x)} &= \beta\langle E(s,x)\rangle_{p(s|x)p(x)} + \sum\sum p(s|x)p(x)\log p(s|x) && (2.106)\\ &= \beta\langle E(s,x)\rangle_{p(s,x)} + \sum\sum p(s,x)\log p(s|x) && (2.107)\\ &= \beta\langle E(s,x)\rangle_{p(s,x)} - H(s|x). && (2.108)\end{aligned}$$

Hence

$$H(s_t|x_{t+1}) = \beta\left(\langle E(s_t, x_{t+1})\rangle_{p(s_t,x_{t+1})} - \langle F_G[p(s_t|x_{t+1})]\rangle_{p(x_{t+1})}\right) \tag{2.109}$$

and

$$H(s_t|x_t) = \beta \left(\langle E(s_t, x_t) \rangle_{p(s_t,x_t)} - \langle F_G[p(s_t|x_t)] \rangle_{p(x_t)} \right). \tag{2.110}$$

We can then rewrite 2.101 in terms of these thermodynamic differences,

$$\begin{aligned} I[s_t; x_t] - I[s_t; x_{t+1}] = \beta \left(\langle E(s_t, x_{t+1}) \rangle_{p(s_t,x_{t+1})} - \langle F_G[p(s_t|x_{t+1})] \rangle_{p(x_{t+1})} \right) \\ - \beta \left(\langle E(s_t, x_t) \rangle_{p(s_t,x_t)} - \langle F_G[p(s_t|x_t)] \rangle_{p(x_t)} \right), \end{aligned} \tag{2.111}$$

which, by rearranging the terms, becomes

$$\begin{aligned} I[s_t; x_t] - I[s_t; x_{t+1}] = \beta \left(\langle E(s_t, x_{t+1}) \rangle_{p(s_t,x_{t+1})} - \langle E(s_t, x_t) \rangle_{p(s_t,x_t)} \right) \\ - \beta \left(\langle F_G[p(s_t|x_{t+1})] \rangle_{p(x_{t+1})} - \langle F_G[p(s_t|x_t)] \rangle_{p(x_t)} \right). \end{aligned} \tag{2.112}$$

Adopting our definition of work W from 2.72,

$$W = \sum_{t=0}^{\tau-1} E(s_t, x_{t+1}) - E(s_t, x_t), \tag{2.113}$$

and using linearity of expectation, we can simplify 2.112 considerably:

$$\begin{aligned} I[s_t; x_t] - I[s_t; x_{t+1}] = \beta \langle W[x_t \to x_{t+1}] \rangle_{p(s_t,x_{t+1})} \\ - \beta \langle \Delta F_G[x_t \to x_{t+1}] \rangle_{p(x_{t+1},x_t)}. \end{aligned} \tag{2.114}$$

Taking the average of the generalized free energy change over the distribution of states s_t, using linearity of expectation again, and recalling from 2.67 that $W_{\text{diss}} = W - \Delta F_G$, we obtain

$$\begin{aligned} I[s_t; x_t] - I[s_t; x_{t+1}] &= \beta \langle W[x_t \to x_{t+1}] - \Delta F_G[x_t \to x_{t+1}] \rangle_{p(x_{t+1},x_t,s_t)} & (2.115) \\ &= \beta \langle W_{\text{diss}}[x_t \to x_{t+1}] \rangle. & (2.116) \end{aligned}$$

□

The energy dissipated by the system as the protocol moves from $x_t \to x_{t+1}$ is then fundamentally equivalent to the amount of system memory that is not predictive of x_{t+1}.

3

Methods

3.1 Model Description

As we saw at the beginning of Chapter 2, neuron models range from detailed, biophysically-realistic neuron models with high dimensional parameter spaces to simple integrate-and-fire neurons that are more mathematically tractable. Choosing a good neuron model then is largely subjective - in addition to choosing a model that conforms well to biological data, it must be suitable for the kinds of analyses planned.

We choose to work with the adaptive exponential integrate-and-fire neuron, developed by Romain Brette and Wolfram Gerstner in 2005, for a variety of reasons.(44) On the one hand, the model is very accurate; when exposed to identical protocols, the adaptive exponential model generated only 3% extra spikes and missed 4% of spikes when compared with more detailed biophysical model neurons.(44)[1] The model is also highly versatile - for different parameter values, the model reflects a variety of real neuron classes.(33) Another benefit is that, by tuning a single parameter a, the model is capable of generating differing degrees of spike-frequency adaptation, allowing us to study the relationship between spike adaptation and the neuron's dissipation of energy. Lastly, low-power *in silica* implementations of this model have been built,(48) allowing us to later test our predictions of energy dissipation as a function of model parameters.

The sub-threshold dynamics of our 2-dimensional model neuron are described by the

[1] Two spikes are considered identical if they occur within 2 ms of each other.

system of stochastic differential equations

$$C\frac{dV}{dt} = f(V) - w + I \tag{3.1}$$

$$\tau_w\frac{dw}{dt} = a(V - E_L) - w, \tag{3.2}$$

where V is voltage and w is the slow adaptation variable. I is the only stochastic term in the system, and is in fact a stochastic control, as it represents the neuron's current input[1] which we control; since synaptic currents are inherently noisy, the stochastic nature of I is reasonable.(65) The system is driven by our choice of $I(t)$, and so in the terminology of sections 2.4-2.6 this is the protocol $x(t)$.

The constants C, τ_w, and E_L represent the cell's capacitance, adaptation time constant, and leak reversal potential, respectively. The parameter a determines subthreshold spike adaptation, and will be a parameter that we "learn" during our maximization of predictive information subject to fixed memory. In a real neuron, a might reflect the composition of potassium ion channels that tend to the hyperpolarize the membrane and lead to spike-frequency adaptation as discussed in 2.1.1.

The function $f(V)$ determines spiking behavior

$$f(V) = -g_L(V - E_L) + g_L\Delta_T \exp\left(\frac{V - V_T}{\Delta_T}\right), \tag{3.3}$$

where the constants g_L, Δ_T, and V_T represent the leak conductance, the slope factor, and the neuron's spiking threshold, respectively. All of these parameters named above, together with their values in a typical neuron, are summarized in Table 3.1.

In addition to these stochastic differential equations that model the neuron's subthreshold dynamics, we declare that the model neuron has fired an action potential at time t whenever $V(t) > 20$ mV, and reset the system at time t according to

$$V(t) = E_L \tag{3.4}$$

$$w(t) = w(t - \delta t) + b. \tag{3.5}$$

We can see then that while a regulates the neuron's subthreshold adaptation, b determines the spike-triggered adaptation adaptation.

[1]The current input I could represent either the synaptic currents of the neuron or the injected current during a current-clamp electrophysiological experiment

3.1.1 Parameters

Let $\theta = \{a, b, C, g_L, E_L, V_T, \Delta_T, \tau_w\}$ be the parameter space for the model 3.1. The parameters $\theta^* = \{C, g_L, E_L, V_T, \Delta_T, \tau_w\}$ are measurable properties of the neuron and are stationary compared to V, w, and I (i.e. $\text{Var}(\theta^*) << \text{Var}(V), \text{Var}(w), \text{Var}(I)$). The 6-dimensional parameter space θ^* is dependent on the surface area of the cell membrane, the relative distribution of different ion channels in the membrane, the partial pressure of oxygen in the neuron[1], pH, mechanical stress on the neuron, and the presence and concentration of G proteins[2]. Since these factors are not relevant to our analysis, we take typical values of θ^* from the neuroscience literature (see Table 3.1) and treat θ^* as fixed.

Parameter	Description	Value
C	membrane capacitance	281 pF
g_L	leak conductance	30 nS
E_L	leak reversal potential	-70.6 mV
V_T	spike threshold	-50.4 mV
Δ_T	slope factor	2 mV
τ_w	adaptation time constant	144 ms
a	subthreshold adaptation	4 nS
b	spike-triggered adaptation	0.0805 nA

Table 3.1: Typical Parameter Values from (44)

.

3.2 Choice of Protocol $x(t)$

Since the code used by neurons is highly adapted to the statistics of the currents that drive the neuron *in vivo*, our choice of protocol will need to be biologically relevant.(36) Our investigation explores how neurons might maximize predictive information with respect to fixed memory; therefore our protocol will need to have nonzero information

[1] This is essentially the amount of O_2 in the fixed volume of the neuron

[2] G proteins are essentially molecular switches that regulate ion channels, enzymes, and other cell signaling cascades. G proteins can be affected by hormones, neurotransmitters, and other signaling factors as well.

in order to arrive at any conclusions. With this in mind, it is interesting to consider the convention in electrophysiology of injecting a simple current step with the intention of teasing out the neuron's transfer function - what amount of information can the neuron encode in this case?

3.2.1 Step Function

Since the step function is such a traditional tool of electrophysiologists, we will first explore the neuron's response to this simple protocol, primarily to confirm our intuition about what I_{mem} and I_{pred} might look like for such a process.

We assume that there are two sources of noise in this protocol: noise in the time when the current steps on and noise in the current step's amplitude. We then define the current to be

$$I(t) = \begin{cases} 0 & \text{if } t < T_{\text{on}} \\ k + \sigma_1 \xi & \text{if } t \geq T_{\text{on}} \end{cases}, \tag{3.6}$$

where $k \in \mathbb{R}$ is some constant current, $\xi \sim \mathcal{N}(0,1)$, and $T_{\text{on}} \sim \mathcal{N}(\mu, \sigma_2^2)$.

We show the neuron spiking in response to this choice of protocol in Figure 1 below.

3.2.2 Ornstein-Uhlenbeck Process

Real neurons in the brain are exposed to thousands of seemingly erratic inputs, which can induce dramatic variability in the firing rate of neurons. This *in vivo*-like background synaptic input current is typically simulated as an Ornstein-Uhlenbeck process, the solution to the Langevin equation

$$dI = -\frac{(I-\mu)}{\tau} dt + \sqrt{D}\, dW_t, \tag{3.7}$$

where τ is the time constant, D is the amplitude of the stochastic component, μ is the mean of the process, and $D\tau/2$ is the variance of the process.(66, 67, 68) Solving this stochastic differential equation gives us the following theorem.

Theorem 3.2.1. *The Langevin equation 3.7 is solved by the Ornstein-Uhlenbeck process*

$$I(t) = \mu + e^{-\frac{(t-s)}{\tau}}(I(s) - \mu) + \sqrt{D} \int_s^t e^{-\frac{(t-u)}{\tau}} dW_u, \tag{3.8}$$

where the integral on the right hand side is an Ito integral.

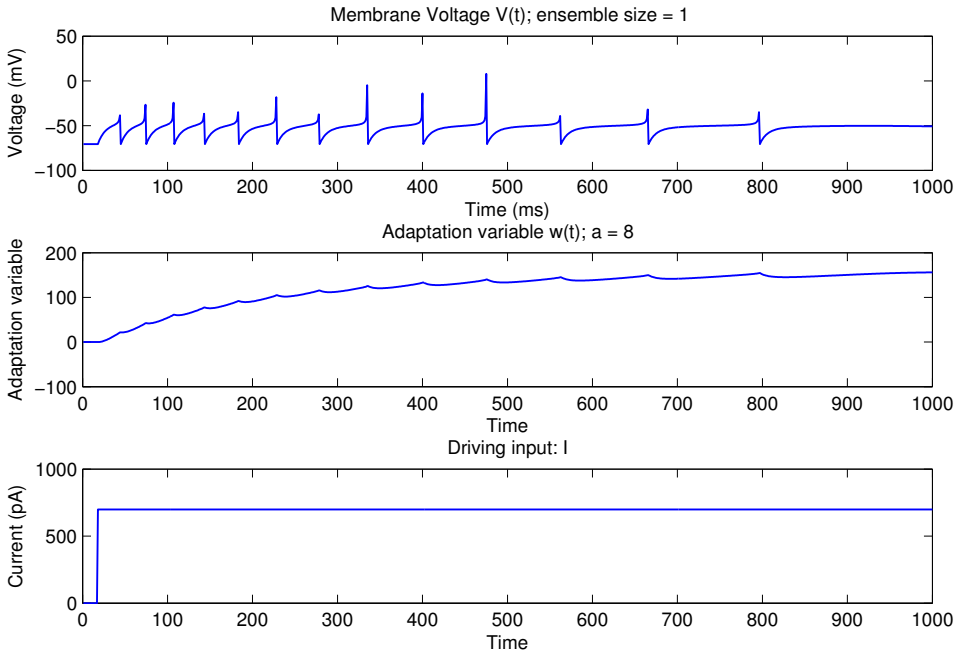

Figure 3.1: Runge-Kutta Simulation (see 3.6) of a single neuron with adaptation parameter $a = 8$ and $I = k + \sigma_1\eta(t)$ where $k = 700$ pA, $\sigma_1 = 1$, $\sigma_2 = 5$ (see 3.2.1), and $\eta \sim \mathcal{N}(0, 1)$.

Proof. In order to subtract off the mean of the process, without loss of generality let $x = I - \mu$; the Langevin equation now becomes $dx = -\frac{x}{\tau}dt + \sqrt{D}dW_t$. Now make the change of variables $y = e^{t/\tau}x$. By the product rule of Ito integrals[1] and the observation that $e^{t/\tau}$ has no stochastic component,

$$\begin{aligned} dy = d(e^{t/\tau} \cdot x) = d(e^{t/\tau} \cdot x) &= e^{t/\tau}dx + xde^{t/\tau} && (3.9)\\ &= e^{t/\tau}dx + \frac{xe^{t/\tau}}{\tau}dt && (3.10)\\ &= e^{t/\tau}\left(-\frac{x}{\tau}dt + \sqrt{D}dW_t\right) + \frac{xe^{t/\tau}}{\tau}dt && (3.11)\\ &= -\frac{xe^{t/\tau}}{\tau}dt + e^{t/\tau}\sqrt{D}dW_t + \frac{xe^{t/\tau}}{\tau}dt && (3.12)\\ &= e^{t/\tau}\sqrt{D}dW_t. && (3.13) \end{aligned}$$

[1]For two stochastic processes $X_t = \mu_1 dt + \sigma_1 dW$ and $Y_t = \mu_2 dt + \sigma_2 dW$, $d(X_t \cdot Y_t) = X_t dY_t + Y_t dX_t + \sigma_1\sigma_2 dt$.(69)

Integrating both sides, we arrive at

$$y(t) = y(s) + \sqrt{D}\int_s^t e^{u/\tau} dW_u. \tag{3.14}$$

By substituting $x(t) = e^{-t/\tau}y$ and then $I(t) = x(t) + \mu$, we find that

$$\begin{aligned} x(t) &= e^{-t/\tau}y(s) + e^{-t/\tau}\sqrt{D}\int_s^t e^{u/\tau} dW_u & (3.15)\\ &= e^{-\frac{(t-s)}{\tau}}x(s) + e^{-t/\tau}\sqrt{D}\int_s^t e^{u/\tau} dW_u & (3.16) \end{aligned}$$

and

$$I(t) = \mu + e^{-\frac{(t-s)}{\tau}}(I(s) - \mu) + \sqrt{D}\int_s^t e^{-\frac{(t-u)}{\tau}} dW_u. \tag{3.17}$$

□

In practice however, background synaptic current is typically described by the standard diffusion approximation to the Ornstein-Uhlenbeck process,

$$I(t) = \mu + e^{-t/\tau}(I_0 - \mu) + \frac{D\tau}{2}(1 - e^{-2t/\tau})\,\eta(t) \tag{3.18}$$

which simply constructs a new stochastic process with the time-dependent mean $\mu + e^{-t/\tau}(I_0 - \mu)$ and variance $(D\tau/2)(1 - e^{-2t/\tau})$ of the original Ornstein-Uhlenbeck process. (70) As usual, $\eta(t)$ is normally distributed with zero mean and unit variance. Note also that, while μ in 3.18 is a constant, we could also add a time dependence to μ such that the mean of the process evolves with some structure $\mu(t)$.

Some authors further simplify the Ornstein-Uhlenbeck by constructing a process similar to 3.18 but with stationary variance $D\tau/2$. If we were to use biologically reasonable parameters for this flavor of approximation and add time dependence to μ, our background synaptic current would be

$$I(t) = g_L\mu(t) + \sqrt{Cg_L}\,\eta(t), \tag{3.19}$$

where g_L is the leak conductance, C is the membrane capacitance, and $\eta(t) \sim \mathcal{N}(0, 1)$.(71) Oftentimes $\mu(t)$ is taken to be sinusoidal, say of the form $\mu(t) = \mu_0 + \mu_1 \cos(2\pi f t)$, where f is the desired frequency of the input. Note however that in this case the variance is stationary, and so there is no real diffusion happening. In fact, if $\mu(t)$ were to lose its

time dependence, we would be reduced to the case of our step function, with $k = g_L\mu$ and $\sigma_1 = \sqrt{Cg_L}$.

For our implementation of the Ornstein-Uhlenbeck process, we will use 3.18 with $\tau = g_L/\sqrt{C}$ and $D = C/\sqrt{g_L}$, such that at $t = 0$ our implementation has equal variance to 3.19. Three sample Ornstein-Uhlenbeck processes are given in Figure 3.2.

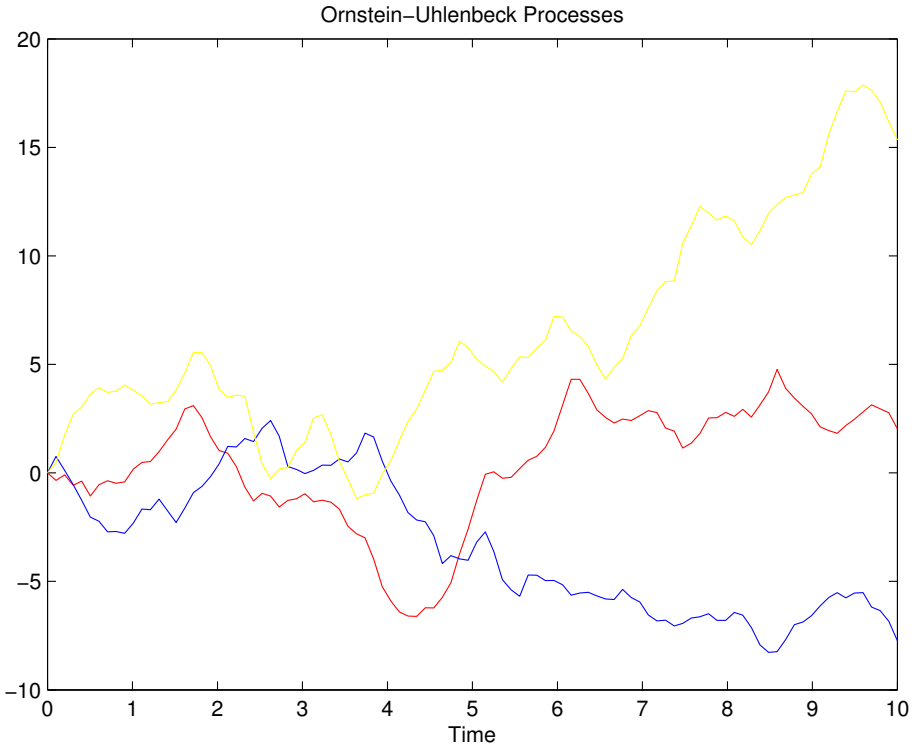

Figure 3.2: Three Ornstein-Uhlenbeck processes all with $\mu = 0$, $\tau = 1/5$, and $D = 5$ (see 3.18) over 10 seconds with $dt = 1/10$.

3.3 Choice of State $s(t)$

We assume that the voltage captures all of the information available to the neuron. Although it is tempting to consider $[V(t); w(t)]$ as the state vector of the system, it is ambiguous how we might define the energy of $w(t)$, which has no real biological

counterpart outside of $V(t)$. As a result, we let the system state be $s(t) = V(t)$.[1]

3.4 Hamiltonian

In treating our model neuron as a physical system, there are several issues we need to address before we can expect to understand the neuron's thermodynamics of prediction. For one, the neuron must be surrounded by a heat bath such that temperature is well-defined and any dissipated heat immediately exits the system. In reality this may not be such a wild notion, since neurons are suspended in matrices of intercellular fluid, blood vasculature, and surrounding tissue.

Even more importantly, we need to define what it means for the neuron's state $s(t)$ to be in thermodynamic equilibrium. In order for us to apply the thermodynamics of prediction, we must have a system that is driven from its equilibrium by our choice of protocol $I(t)$. In the absence of this protocol (or equivalently, for fixed $I(t) = 0$), the equilibrium distribution of system states is given by the Boltzmann distribution

$$p_{\text{eq}}(s) = \frac{1}{Z} e^{-\beta E_s}, \tag{3.20}$$

where E_s is the total energy of the system in state s and Z is the partition function

$$Z = \sum_s e^{-\beta E_s}. \tag{3.21}$$

Let the protocol in general be $x(t)$. For certain physical systems, we can write the total energy E_s for fixed x as a smooth, scalar function $H(s, x, t)$. This $H(s, x, t)$ is called the Hamiltonian of the system. In general, a Hamiltonian system is a dynamical system of $2n$, first order, ordinary differential equations

$$\dot{z} = J\nabla H(q, p, x, t), \qquad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \qquad \nabla H(q, p, x, t) = \begin{pmatrix} \frac{\partial H}{\partial q} \\ \frac{\partial H}{\partial p} \end{pmatrix}, \tag{3.22}$$

where I is the usual identity matrix. To see the relationship between $H(s, x, t)$ and $H(q, p, x, t)$ we need only set the system's state vector $s = \{q, p\}$. We also note that in general the control protocol x is optional; even without this control the system would

[1] In future work, it might be interesting to consider instantaneous firing rate $v(t)$ as the state of the system, and see what fraction of information $v(t)$ captures from $V(t)$.

still be Hamiltonian.

Although in physics the Hamiltonian represents the total energy of the system with canonical coordinates q and momenta p, there is a history of state space models in neuroscience as well; however, the Hamiltonian in these cases never designates the total energy of the system, but rather contains "pseudoenergy" terms.(72, 73, 74, 75, 76) In contrast, our Hamilton must represent the total energy of the neuron - otherwise our calculation of $\beta\langle W_{\text{diss}}\rangle$ looses its thermodynamic meaning.

Fortunately, our model neuron can be represented by a fairly simple RC circuit, with the resistors and capacitors in parallel. Without loss of generality, we will compute the model neuron's thermodynamic equilibrium at input current $I(t) = 0$. Since our neuron at equilibrium is in steady state, we must have

$$a(V - E_L) - w = \tau_w \frac{dw}{dt} = 0, \tag{3.23}$$

which must be true for all values of $a \in \mathbb{R}$. At this point we restrict ourselves to calculating the equilibrium at time $t = 0$ before any adaptation occurs; i.e., adaptation $w = 0$. These conditions leave us no choice but for the steady state voltage to be $V = E_L$.

In the absence of input current, there must be no current leaving the circuit by Kirchoff's law, which states that the sum of currents flowing into and out of a given node must be zero. Then no current is flowing across the resistor or capacitor, and all of the energy in the circuit must be stored as stationary charge in the capacitor. The energy stored in the capacitor is equal to the energy needed (or equivalently, the work done) to charge it. If the capacitor holds a positive charge $+q$ on one side and $-q$ on the other, then moving a small charge dq from one side to the other against the potential difference $V = q/C$ (2.1.1) requires energy dE,

$$dE = V dq = \frac{q}{C} dq, \tag{3.24}$$

which we can then integrate over the entire charge Q on the capacitor,

$$E = \int_0^Q \frac{q}{C} dq = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2} C V^2. \tag{3.25}$$

Since our voltage has steady state E_L, with a simple change of variables the total energy E stored on our membrane capacitor is then

$$E(V) = \frac{1}{2}C(V - E_L)^2. \tag{3.26}$$

Substituting our total energy in 3.20 we then have the equilibrium distribution of states

$$p_{\text{eq}}(V) = \frac{1}{Z}e^{-\frac{\beta C(V-E_L)^2}{2}} = \frac{1}{Z}e^{-\frac{(V-E_L)^2}{2(1/\beta C)}}. \tag{3.27}$$

Noting the similarity this bears to the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3.28}$$

we conclude that $p_{\text{eq}}(V) \sim \mathcal{N}(E_L, 1/\beta C)$. For our implementation, we will take thermodynamic $\beta = \frac{1}{k_B T}$, where k_B is the Boltzmann constant, to have temperature $T = 310.65$ K, the average core human body temperature, and capacitance $C = 281$ pF as per Table 3.1.

3.5 From ODEs to SDEs

We can write stochastic differential equations as the Langevin equation

$$dX(t) = A(X(t))\,dt + B(X(t))\,dW(t) \tag{3.29}$$

$$X(t_0) = x_0. \tag{3.30}$$

In the case where $X(t)$ is scalar, 3.29 has the general solution

$$X(t) = X(t_0) + \int_{t_0}^{t} A(X(s), s)\,ds + \int_{t_0}^{t} B(X(s), s)\,dW(s), \tag{3.31}$$

where the second integral is the Ito integral. In this manner, we will now formulate our complete system (encapsulating both the system's state and the driving protocol) as a canonical system of stochastic differential equations.

3.5.1 Current Step Protocol

Suppose current I is a simple noisy current step $I(t) = k + \sigma_1\xi$ as in 3.2.1 where $k, \sigma_1 \in \mathbb{R}$ are constants and $\xi \sim \mathcal{N}(0, 1)$. We will only treat this part of the piecewise function, as the $I(t) = 0$ case follows trivially by setting $k = \sigma_1 = 0$.

Letting

$$\mathbf{Z}(t) = \begin{bmatrix} V(t) \\ w(t) \end{bmatrix}, \qquad e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{3.32}$$

we can express our system of equations 3.1 in the form specified by 3.29 by considering the vectors

$$\mathbf{A}(\mathbf{Z}(t)) = \begin{bmatrix} \frac{1}{C}\left(-g_L e_1^T \mathbf{Z}(t) + g_L E_L + g_L \Delta_T \exp\left(\frac{e_1^T \mathbf{Z}(t) - V_T}{\Delta_T}\right) - e_2^T \mathbf{Z}(t) + k\right) \\ \frac{1}{\tau_w}\left(a(e_1^T \mathbf{Z}(t) - E_L) - e_2^T \mathbf{Z}(t)\right) \end{bmatrix}, \tag{3.33}$$

and

$$\mathbf{B}(\mathbf{Z}(t)) = \begin{bmatrix} \sigma_1/C \\ 0 \end{bmatrix}. \tag{3.34}$$

With $\mathbf{A}$ and $\mathbf{B}$ specified, our system of stochastic differential equations can then be stated in the canonical vector form

$$d\mathbf{Z}(t) = \mathbf{A}(\mathbf{Z}(t))\, dt + \mathbf{B}(\mathbf{Z}(t))\, d\mathbf{W}(t). \tag{3.35}$$

Although our system is not scalar, we can adapt 3.31 for our purposes here as

$$\mathbf{Z}(t) = \mathbf{Z}(t_0) + \int_{t_0}^{t} \mathbf{A}(\mathbf{Z}(s), s)\, ds + \int_{t_0}^{t} \mathbf{B}(\mathbf{Z}(s), s)\, d\mathbf{W}(s), \tag{3.36}$$

where now $\mathbf{W}(s)$ is a 2-dimensional Wiener process.

3.5.2 Ornstein-Uhlenbeck Process Protocol

We now suppose that $I(t) = g_L\mu(t) + \sqrt{Cg_L}\,\eta(t)$ as in 3.19. Letting $I(0) = 0$ and recycling $\mathbf{Z}(t)$ from above, our vectors $\mathbf{A}$ and $\mathbf{B}$ become

$$\mathbf{A}(\mathbf{Z}(t)) = \begin{bmatrix} \frac{1}{C}\left(-g_L e_1^T \mathbf{Z}(t) + g_L E_L + g_L \Delta_T \exp\left(\frac{e_1^T \mathbf{Z}(t) - V_T}{\Delta_T}\right) - e_2^T \mathbf{Z}(t) + \mu - \mu e^{-t/\tau}\right) \\ \frac{1}{\tau_w}\left(a(e_1^T \mathbf{Z}(t) - E_L) - e_2^T \mathbf{Z}(t)\right) \end{bmatrix}, \tag{3.37}$$

and

$$\mathbf{B}(\mathbf{Z}(t)) = \begin{bmatrix} \frac{D\tau}{2C}(1 - e^{-2t/\tau}) \\ 0 \end{bmatrix}, \tag{3.38}$$

with the canonical equations 3.35 and 3.36 identical to the ones above.

3.6 Numerical Methods

Unfortunately, very few systems of stochastic differential equations are solvable - in fact virtually any system that can be solved analytically concerns the case when the random process X is scalar.(65) In the case of nonlinear stochastic differential equations, explicit solutions are unobtainable and we must resort to numerical methods, which can be difficult in and of themselves.(77) In choosing a numerical method, we seek to find an approximate numerical solution $\mathbf{Y}(t)$ to the true solution $\mathbf{Z}(t)$ of 3.35 with a certain degree of accuracy. Although a number of methods exist for numerically approximating stochastic differential equations, their accuracies can be compared using the notion of strong convergence. A numerical method is said to have *strong convergence* equal to γ if there exists a constant C such that

$$\langle |Z_i(t) - X_i(t)| \rangle \leq C\Delta t^{\gamma} \tag{3.39}$$

for any fixed $\tau = n\Delta t \in [0, T]$, Δt sufficiently small, and for any $1 \leq i \leq M$, where $\mathbf{Z}$ and $\mathbf{Y}$ are M-dimensional stochastic processes.(78) With this criterion in mind, we use the stochastic Runge-Kutta method as its order of strong convergence is never less than 3,[1] and generalizes well to dimensions > 1.(79)

3.6.1 Time step δt

In our system, $\mathbf{Z}(t)$ is a two dimensional vector with elements $V(t)$ and $w(t)$. After discretizing 3.1, suppose we first compute $V(t + \delta t)$ and then $w(t + \delta t)$. Both $V(t + \delta t)$ and $w(t + \delta t)$ are properly functions of V and w, and so when it is time to compute $w(t + \delta t)$, we must choose whether it is a function of $V(t)$ or $V(t + \delta t)$. In simulating most dynamical systems, the time step δt must be small enough such that this choice does not matter.[2] However, we find that even choosing $\delta t = 1$ is small enough since the factors $1/C$ and $1/\tau_w$ are sufficiently small. Furthermore, simulating the system with $w(t + \delta t)$ as a function of $V(t + \delta t)$ fails to produce the spiking behavior of the model, since the action potential is of such short duration.

[1] And the quadratic mean error of the approximate solution is never greater than $\mathcal{O}(\Delta t^3)$.
[2] Personal communication with Dr. Robert Shaw.

3.6.2 Runge-Kutta Method

Suppose we want to find the numerical solution $\mathbf{Y}(t)$ over the time interval $[0, T]$. First we discretize our system by partitioning $[0, T]$ into N equal sub-intervals of width $\delta = T/N > 0$,

$$0 = \tau_0 < \tau_1 < \cdots < \tau_N = T. \tag{3.40}$$

Following (80), we then set $\mathbf{Y}(0) = \mathbf{Z}(0)$ and define the Runge-Kutta solution $\mathbf{Y}(t)$ recursively for $1 \leq n \leq N$ as

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{A}(\mathbf{Y}_n)\delta + \mathbf{B}(\mathbf{Y}_n)\Delta\mathbf{W}_n + \frac{1}{2}\left(\mathbf{B}(\hat{\mathbf{Y}}_n) - \mathbf{B}(\mathbf{Y}_n)\right)\left((\Delta\mathbf{W}_n)^2 - \delta\right)\delta^{-1/2}, \tag{3.41}$$

where

$$\Delta\mathbf{W}_n = \mathbf{W}_{\tau_{n+1}} - \mathbf{W}_{\tau_n} \tag{3.42}$$

are independent and identically distributed normal random variables $\sim \mathcal{N}(0, \delta)$ and

$$\hat{\mathbf{Y}}_n = \mathbf{Y}_n + \mathbf{A}(\mathbf{Y}_n)\delta + \mathbf{B}(\mathbf{Y}_n)\delta^{1/2}. \tag{3.43}$$

Simplifications

When the current $I(t)$ is a simple noisy current step $k + \sigma_1\xi$, $\mathbf{B}$ (see 3.34) is not a function of $\mathbf{Y}_n$ - in fact, it is just a constant vector. Then $\mathbf{B}(\hat{\mathbf{Y}}_n) = \mathbf{B}(\mathbf{Y}_n)$ and 3.41 becomes

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{A}(\mathbf{Y}_n)\delta + \mathbf{B}(\mathbf{Y}_n)\Delta\mathbf{W}_n, \tag{3.44}$$

which is just a simple generalization of the Euler method for ordinary differential equations to stochastic differential equations called the Euler-Maruyama method.(78) Here $\mathbf{A}$ and $\mathbf{B}$ are as in 3.33 and 3.34.

Alternatively, when the current $I(t)$ is the Ornstein-Uhlenbeck process given in 3.18, $\mathbf{B}(\mathbf{Y}_n)$ and $\mathbf{B}(\hat{\mathbf{Y}}_n)$ are not constant vectors, but are rather time-dependent. They are not, however, functions of $\mathbf{Y}_n$, and so as with the current step, $\mathbf{B}(\hat{\mathbf{Y}}_n) = \mathbf{B}(\mathbf{Y}_n)$ and the Runge-Kutta method reduces to the Euler-Maruyama method.

3.7 Transition Probabilities

Now that we have clarified the state $s(t)$ and protocol $x(t)$ of the system, we need to determine the joint and marginal probability distributions of $s(t)$ and $x(t)$ through time. Although we find these transition probabilities empirically, analytically we can examine the evolution of probability distributions via the Fokker-Planck equation.

3.7.1 Markov Property

Looking back at the approximation 3.41 of our system, we can notice that each $\mathbf{Y}_{n+1}$ is dependent on only the previous $\mathbf{Y}_n$ and not on any previous histories. On account of this, we can say that our system obeys the Markov property

$$\mathbb{P}(\mathbf{Y}_n|\mathbf{Y}_{n-1}, \mathbf{Y}_{n-2}, \ldots, \mathbf{Y}_0) = \mathbb{P}(\mathbf{Y}_n|\mathbf{Y}_{n-1}). \tag{3.45}$$

In this paper, we calculate all of our probability distributions empirically using ensembles of 10^3 neurons.

4

Results

4.1 Memory, Predictive Power, and Nonpredictive Information

For each of the protocols 3.2.1 and 3.2.2 we calculated the information theoretic quantities I_{mem}, I_{pred}, and $I_{\text{nonpred}} = I_{\text{mem}} - I_{\text{pred}}$ as defined by (4) and restated in 2.6.

4.1.1 Current Step

After $t > T_{\text{on}}^{(i)}$ for all neurons i, we expect for $I_{\text{mem}} = I_{\text{pred}} = 0$. Using the definition 2.37 of mutual information in terms of conditional entropies, we have

$$I_{\text{mem}} = I[s_t; x_t] = H(x_t) - H(x_t|s_t). \tag{4.1}$$

In the case where the current step only occupies one value k (to the accuracy of our empirical probability distribution), we have $p(k) = p(k|s_t) = 1$. Using 2.23 and 2.24 we then have

$$\begin{aligned} I_{\text{mem}} &= H(x_t) - H(x_t|s_t) && (4.2)\\ &= -\sum p(x_t) \log p(x_t) - \sum\sum p(x_t, s_t) \log p(x_t|s_t) && (4.3)\\ &= -p(k) \log p(k) - \sum p(k, s_t) \log p(k|s_t) && (4.4)\\ &= -\log 1 - \sum \log 1 && (4.5)\\ &= 0, && (4.6) \end{aligned}$$

where $0 \cdot \log 0 := 0$. Substituting $x_{t+1} = x_t$, it follows that $I_{\text{pred}} = 0$ as well.

Testing this empirically confirmed our suspicion; once the protocols $I(t) = k$, for fixed $k \in \mathbb{R}$, we measured $I_{\text{mem}} = I_{\text{pred}} = 0$.

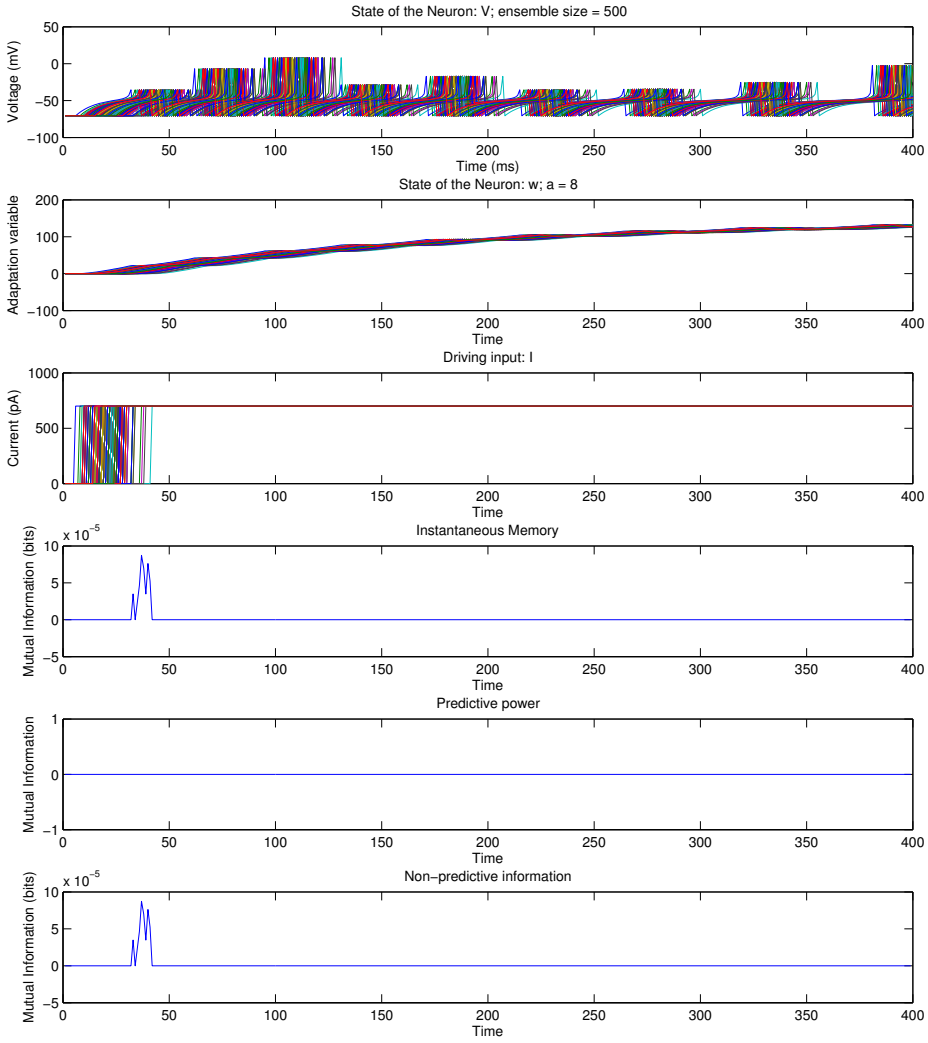

Figure 4.1: Runge-Kutta Simulation of 500 neurons with adaptation parameter $a = 8$ and $I = k + \sigma_1 \eta(t)$ where $k = 700$ pA, $\sigma_1 = 1$, $\sigma_2 = 5$ (see 3.2.1), and $\eta \sim \mathcal{N}(0, 1)$ The value of k was chosen to reflect the minimal current at which the neuron spiked.

The peak of I_{mem} around 40 ms shrunk with increasing sample size, and so we at-

tribute the small amount of mutual information $\approx 10^{-4}$ bits to finite sampling errors. We also investigated the effect changing the adaptation parameter a had on nonpredictive information (see Figure 4.2). It remains to be determined that the fluctuations seen in Figure 4.2 are indeed statistically insignificant.

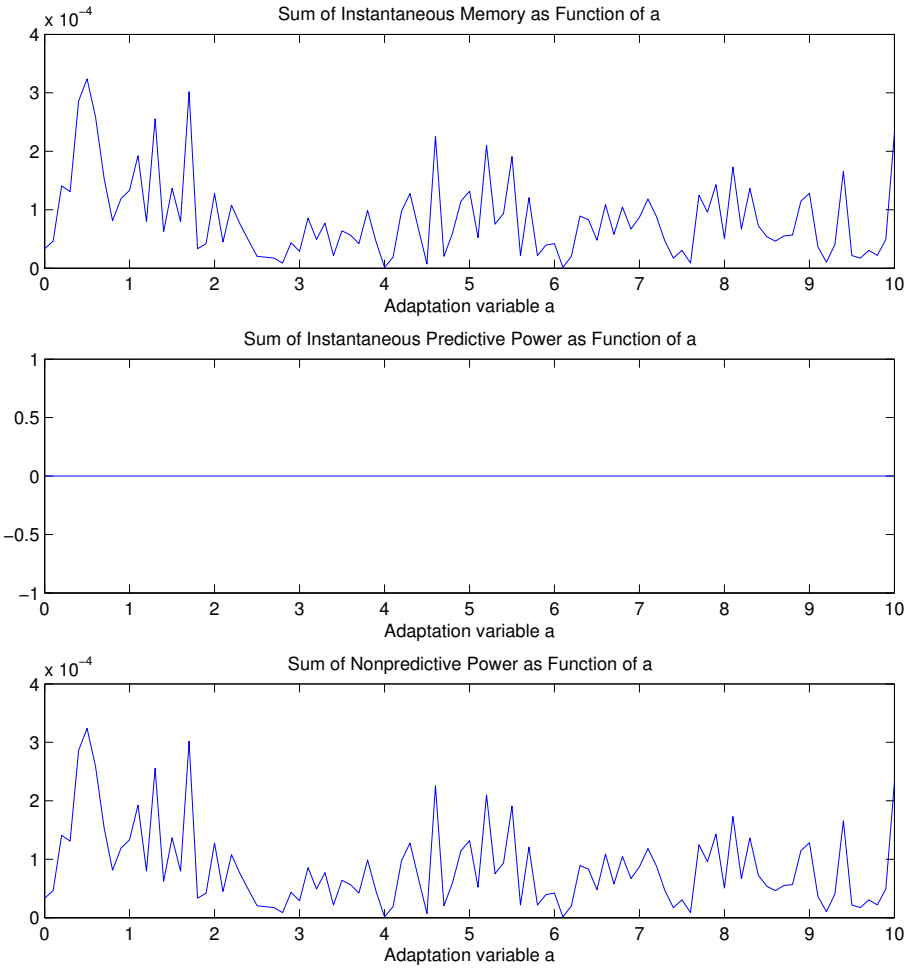

Figure 4.2: Memory, predictive power, and nonpredictive information as a function of a under the current step regime. The information theoretic values are calculated using Runge-Kutta simulations of 1,000 neurons for each of 100 values of $a \in [0, 10]$.

4.1.2 Ornstein-Uhlenbeck Process

Next we substituted the current step - an artificial, zero information process - to the Ornstein-Uhlenbeck process that resembles background synaptic activity (see 3.2.2). The first thing we notice is the variety of neuronal responses (see Figures 4.3-4.5).

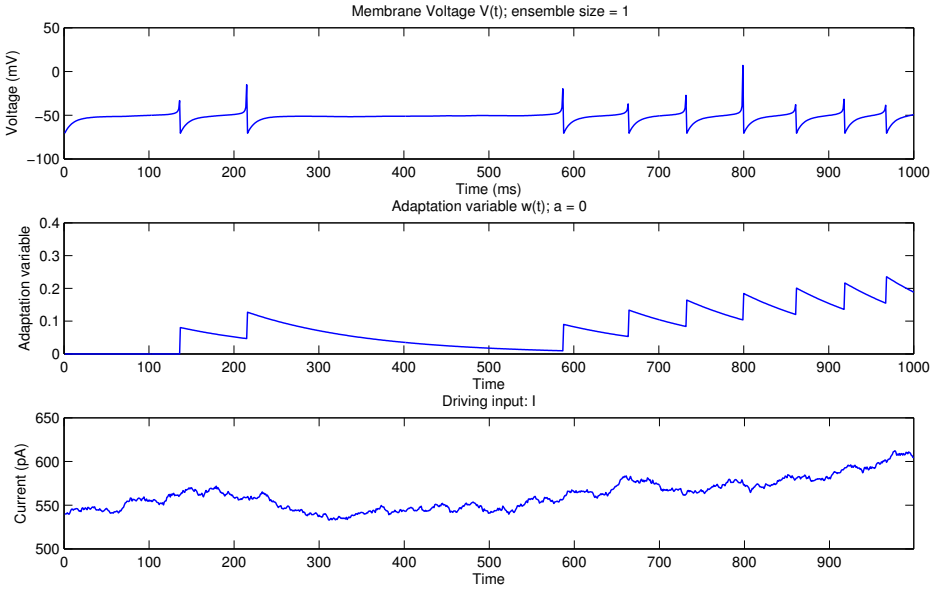

Figure 4.3: Runge-Kutta simulation of a single neuron (with no adaptation, $a = 0$) under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA, $\tau = 1/5$, and $D = 5$.

We also noticed that adaptation had a much stronger influence, suppressing all neuron spiking when $a > 6$ for the same μ values that led to robust spiking in non-adapting cells (compare figures 4.6 and 4.7).

Lastly, we computed the instantaneous memory $I_{\text{mem}}(t)$, predictive power $I_{\text{pred}}(t)$, and nonpredictive information $I_{\text{mem}}(t) - I_{\text{pred}}(t)$ and found that over time, instantaneous nonpredictive information converged to zero. However - and this is startling - the decrease in nonpredictive information is accompanied by a decrease in memory and predictive power in the theoretical neuron without adaptation, but *not* in the neuron with adaptation (Figures 4.8 and 4.9). Neurons with adaptation are able to maintain their level of instantaneous memory and predictive power while decreasing their nonpredictive information; combining this with the results from (4), this implies that

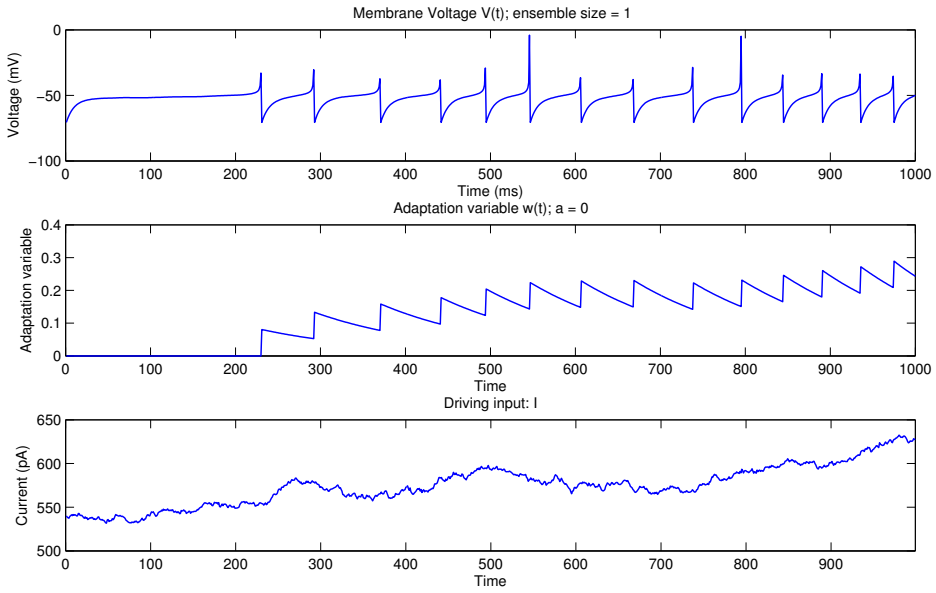

Figure 4.4: Runge-Kutta simulation of a single neuron (with no adaptation, $a = 0$) under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA, $\tau = 1/5$, and $D = 5$.

neurons with adaptation are able to reduce their energy dissipation at fixed memory. Happily, we see a clear minimum in Figure 4.10, and so we can plan on minimizing our neuron's nonpredictive information subject to fixed memory by optimizing adaptation a. We can then test this optimization against experimental measurements of $\beta\langle W_{\text{diss}}\rangle$ *in silica*.

4.2 Energy Dissipation

Using our optimal adaptation value $a = 9$ (see Figure 4.10), we can then calculate the lower bound of energy dissipation in our adaptive exponential integrate-and-fire neuron by applying the relationship

$$\beta\langle W_{\text{diss}}\rangle \geq I_{\text{mem}} - I_{\text{pred}} \tag{4.7}$$

from (4). Over the 1 second protocol length, the system accumulated $I_{\text{mem}} - I_{\text{pred}} = 13.1749$ bits of nonpredictive information. Using the equality 4.7, $T = 310.65$ K, and Boltzmann constant $k_B = 1.3806503 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}, we find that the average

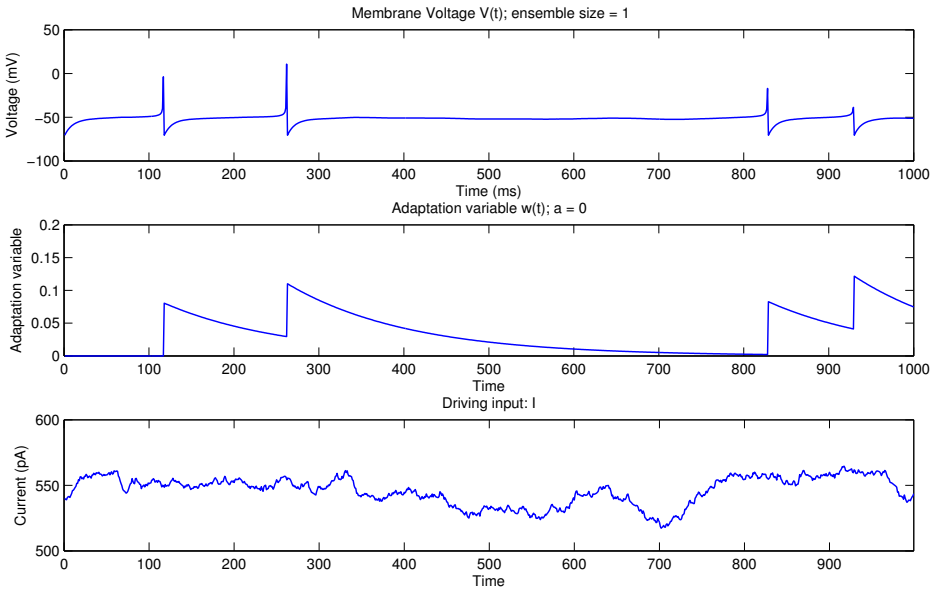

Figure 4.5: Runge-Kutta simulation of a single neuron (with no adaptation, $a = 0$) under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA, $\tau = 1/5$, and $D = 5$.

energy dissipated is

$$\begin{aligned} \langle W_{\text{diss}} \rangle &= k_B T \left(I_{\text{mem}} - I_{\text{pred}} \right) && (4.8) \\ &= 13.1749 k_B T && (4.9) \\ &= 5.65070164 \times 10^{-20} \text{ J}. && (4.10) \end{aligned}$$

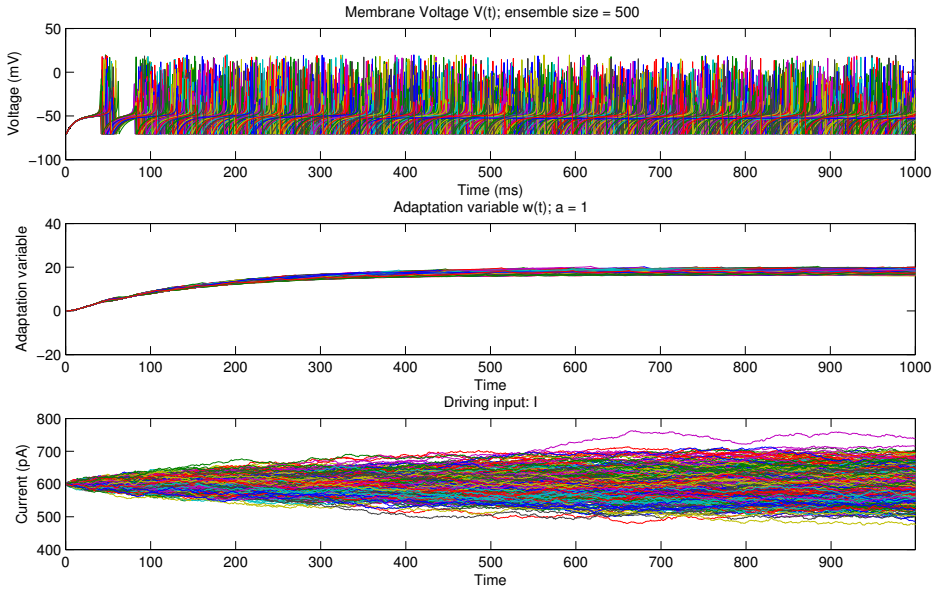

Figure 4.6: Runge-Kutta simulation of 500 neurons under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA and $a = 0$.

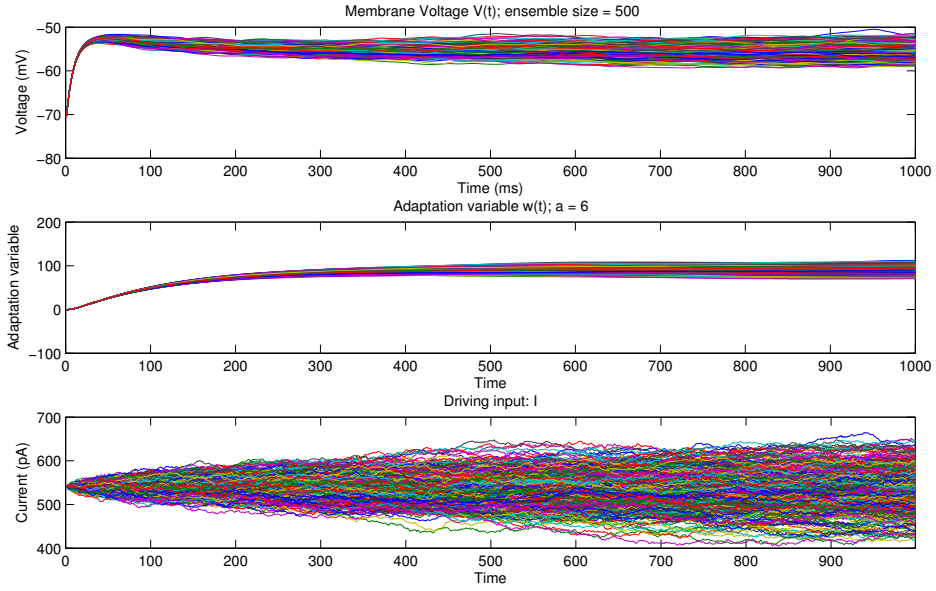

Figure 4.7: Runge-Kutta simulation of 500 neurons under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA and $a = 6$.

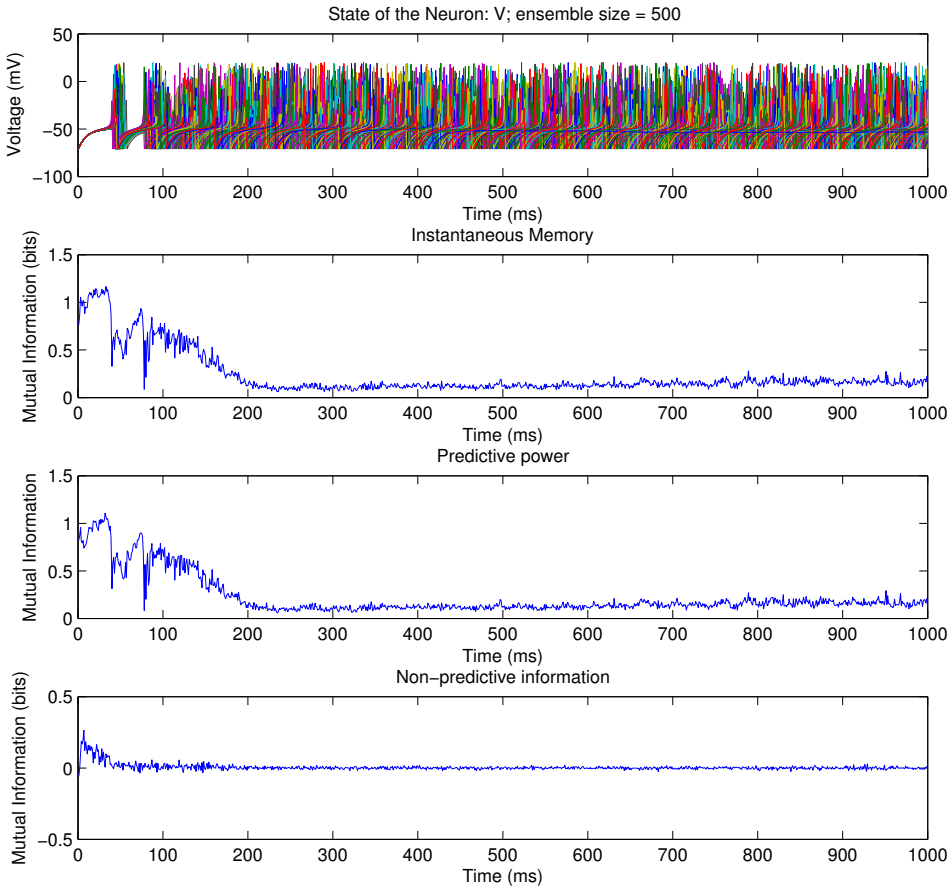

Figure 4.8: Runge-Kutta simulation of 500 neurons under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA and $a = 0$.

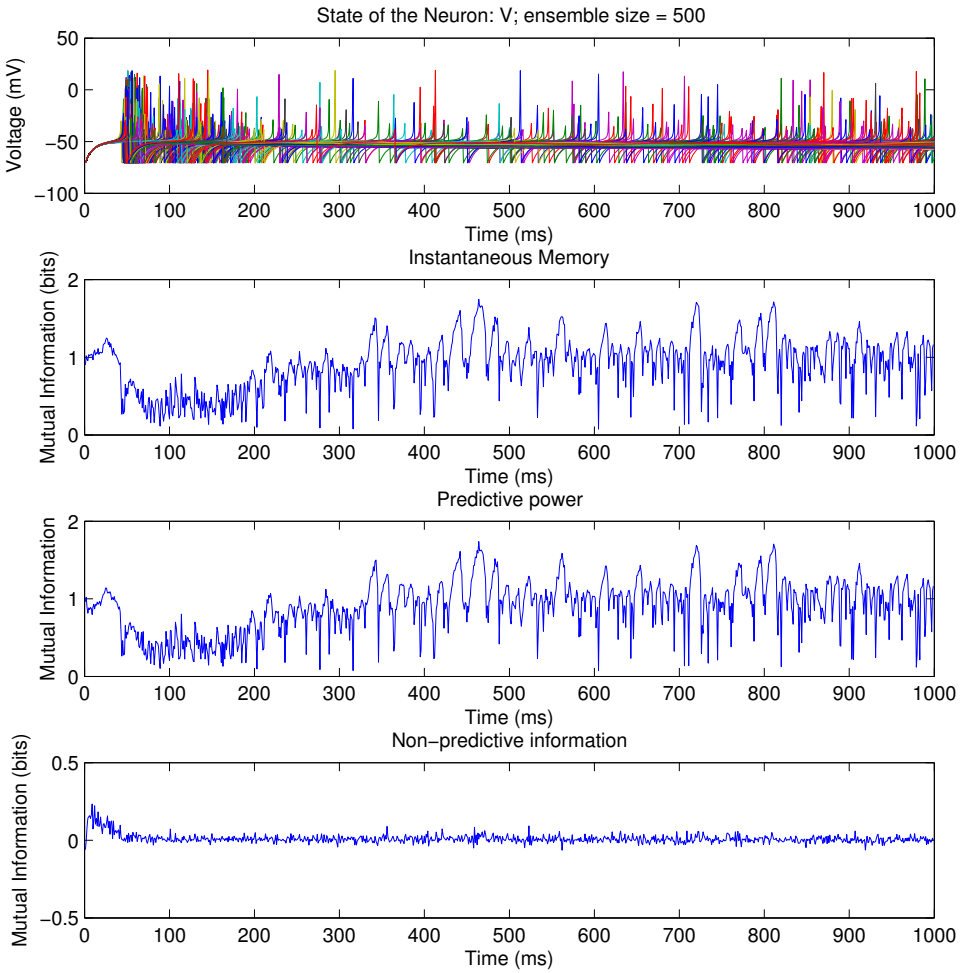

Figure 4.9: Runge-Kutta simulation of 500 neurons under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA and $a = 6$.

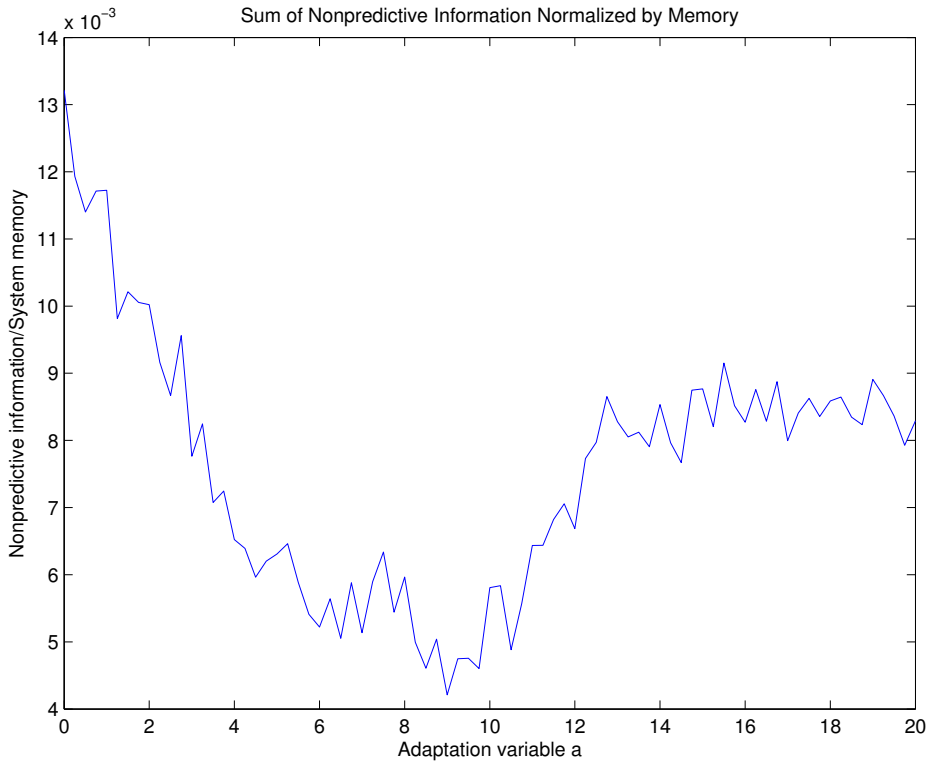

Figure 4.10: Nonpredictive information I_{nonpred} normalized by the system's memory as a function of adaptation a (80 samples). Each sample represents a Runge-Kutta simulation of 500 neurons under the Ornstein-Uhlenbeck protocol with $\mu = 540$ pA.

References

[1] R. Landauer. **Computation: A fundamental physical view**. *Physica Scripta*, **35**:88, 1987. 1, 26

[2] S. Lloyd. **Computational capacity of the universe**. *Physical Review Letters*, **88**(23):237901, 2002. 1

[3] G. Neske. **The notion of computation is fundamental to an autonomous neuroscience**. *Complexity*, **16**(1):10–19, 2010. 1

[4] Susanne Still, David A. Sivak, Anthony J. Bell, and Gavin E. Crooks. **The thermodynamics of prediction**. 03 2012. 1, 2, 3, 5, 20, 21, 22, 24, 27, 44, 47, 48

[5] E.W. Weisstein. **Damped Simple Harmonic Motion–Critical Damping**. *From MathWorld–A Wolfram Web Resource*, **1**, 2010. 1

[6] F. Rieke. *Spikes: exploring the neural code*. The MIT Press, 1999. 1, 6

[7] S.B. Laughlin, R.R.R. van Steveninck, and J.C. Anderson. **The metabolic cost of neural information**. *Nature neuroscience*, **1**(1):36–41, 1998. 2, 7

[8] G.H. Recanzone, M.M. Merzenich, and C.E. Schreiner. **Changes in the distributed temporal response properties of SI cortical neurons reflect improvements in performance on a temporally based tactile discrimination task**. *Journal of Neurophysiology*, **67**(5):1071–1091, 1992. 2

[9] F. Rieke and DA Baylor. **Single-photon detection by rod cells of the retina**. *Reviews of Modern Physics*, **70**(3):1027, 1998. 2

[10] Eric R Kandel, James H Schwartz, and Thomas M Jessell. *Principles of neural science*. McGraw-Hill, Health Professions Division, New York, 4th ed edition, 2000. 2, 4, 6, 7

[11] Peter Dayan and L. F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Massachusetts Institute of Technology Press, Cambridge, Mass., 2001. 4

[12] R.W. Williams and K. Herrup. **The control of neuron number**. *Annual Review of Neuroscience*, **11**(1):423–453, 1988. 5

[13] D.A. Drachman. **Do we have brain to spare?** *Neurology*, **64**(12):2004–2005, 2005. 5

[14] Philip S. Ulinski. **Fundamentals of Computational Neuroscience**. *Unpublished Manuscript*, 2010. 5, 7, 9

[15] R. Yuste. **Circuit neuroscience: the road ahead**. *Frontiers in neuroscience*, **2**(1):6, 2008. 5

[16] A.J. Bell. **Towards a cross-level theory of neural learning**. In *27th international workshop on Bayesian inference and maximum entropy methods in science and engineering, AIP Conference Proceedings*, **954**, pages 56–73, 2007. 5

[17] S.M. Sherman and RW Guillery. **Functional organization of thalamocortical relays**. *Journal of Neurophysiology*, **76**(3):1367–1395, 1996. 5

[18] G. Laurent and H. Davidowitz. **Encoding of olfactory information with oscillating neural assemblies**. *Science*, **265**(5180):1872–1875, 1994. 5

[19] J. Gautrais and S. Thorpe. **Rate coding versus temporal order coding: a theoretical approach**. *Biosystems*, **48**(1-3):57–65, 1998. 6

[20] R.V. Rullen and S.J. Thorpe. **Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex**. *Neural computation*, **13**(6):1255–1283, 2001. 6

[21] V. Booth and A. Bose. **Neural mechanisms for generating rate and temporal codes in model CA3 pyramidal cells**. *Journal of neurophysiology*, **85**(6):2432–2445, 2001. 6

[22] J. Huxter, N. Burgess, and J. OKeefe. **Independent rate and temporal coding in hippocampal pyramidal cells**. *Nature*, **425**(6960):828, 2003. 6

[23] MR Mehta, AK Lee, and MA Wilson. **Role of experience and oscillations in transforming a rate code into a temporal code**. *Nature*, **417**(6890):741–746, 2002. 6

[24] Bertil Hille. *Ionic channels of excitable membranes*. Sinauer Associates, Sunderland, Mass., 2nd ed edition, 1992. 6, 7, 9, 10

[25] J.W. Mink, R.J. Blumenschine, and D.B. Adams. **Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis**. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **241**(3):R203–R212, 1981. 7

[26] J.E. Niven and S.B. Laughlin. **Energy limitation as a selective pressure on the evolution of sensory systems**. *Journal of Experimental Biology*, **211**(11):1792–1804, 2008. 7, 8

[27] HB Barlow. **Trigger features, adaptation and economy of impulses**. *Information Processing in the Nervous System*, pages 209–230, 1969. 8

[28] W.B. Levy and R.A. Baxter. **Energy efficient neural codes**. *Neural Computation*, **8**(3):531–543, 1996. 8

[29] D.J. Field. **What is the goal of sensory coding?** *Neural computation*, **6**(4):559–601, 1994. 8

[30] Michael S Gazzaniga. *The cognitive neurosciences*. MIT Press, Cambridge, Mass., 4th ed edition, 2009. 8

[31] B.A. Olshausen and D.J. Field. **Sparse coding of sensory inputs**. *Current opinion in neurobiology*, **14**(4):481–487, 2004. 8

[32] A. Hasenstaub, S. Otte, E. Callaway, and T.J. Sejnowski. **Metabolic cost as a unifying principle governing neuronal biophysics**. *Proceedings of the National Academy of Sciences*, **107**(27):12329, 2010. 8

[33] E.M. Izhikevich. **Simple model of spiking neurons**. *Neural Networks, IEEE Transactions on*, **14**(6):1569–1572, 2003. 8, 30

[34] G. Fuhrmann, H. Markram, and M. Tsodyks. **Spike frequency adaptation and neocortical rhythms**. *Journal of neurophysiology*, **88**(2):761–770, 2002. 8, 9

[35] A.L. Fairhall, G.D. Lewen, W. Bialek, and RR de Ruyter van Steveninck. **Multiple timescales of adaptation in a neural code**. *Advances in neural information processing systems*, pages 124–130, 2001. 8

[36] A.L. Fairhall, G.D. Lewen, W. Bialek, and R.R. de Ruyter van Steveninck. **Efficiency and ambiguity in an adaptive neural code**. *Nature*, **412**(6849):787–792, 2001. 8, 32

[37] J. Benda and A.V.M. Herz. **A universal model for spike-frequency adaptation**. *Neural computation*, **15**(11):2523–2564, 2003. 8, 9

[38] I.A. Fleidervish, A. Friedman, and MJ Gutnick. **Slow inactivation of Na+ current and slow cumulative spike adaptation in mouse and guinea-pig neocortical neurones in slices.** *The Journal of physiology*, **493**(Pt 1):83–97, 1996. 9

[39] W. Gerstner and R. Naud. **How good are neuron models?** *Science*, **326**(5951):379–380, 2009. 9

[40] LF Abbott et al. **Lapicques introduction of the integrate-and-fire model neuron (1907)**. *Brain research bulletin*, **50**(5):303–304, 1999. 9

[41] W. Gerstner and W.M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge Univ Pr, 2002. 10

[42] A.L. Hodgkin and A.F. Huxley. **A quantitative description of membrane current and its application to conduction and excitation in nerve**. *Bulletin of mathematical biology*, **52**(1):25–71, 1990. 10

[43] E.M. Izhikevich. *Dynamical systems in neuroscience: the geometry of excitability and bursting*. The MIT press, 2007. 10

[44] Romain Brette and Wulfram Gerstner. **Adaptive exponential integrate-and-fire model as an effective description of neuronal activity**. *J Neurophysiol*, **94**(5):3637–42, Nov 2005. 11, 30, 32

[45] G. Indiveri, B. Linares-Barranco, T.J. Hamilton, A. Van Schaik, R. Etienne-Cummings, T. Delbruck, S.C. Liu, P. Dudek, P. Häfliger, S. Renaud, et al. **Neuromorphic silicon neuron circuits**. *Frontiers in neuroscience*, **5**, 2011. 11

[46] C. Mead. **Neuromorphic electronic systems**. *Proceedings of the IEEE*, **78**(10):1629–1636, 1990. 11

[47] J. Dethier, P. Nuyujukian, C. Eliasmith, T. Stewart, S.A. Elassaad, K.V. Shenoy, and K. Boahen. **A Brain-Machine Interface Operating with a Real-Time Spiking Neural Network Control Algorithm**. *Advances in Neural Information Processing Systems (NIPS) 24*, 2011. 11

[48] G. Indiveri. **A low-power adaptive integrate-and-fire neuron circuit**. In *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on*, **4**, pages IV–820. Ieee, 2003. 11, 30

[49] C. E. Shannon. **A mathematical theory of communication**. *SIGMOBILE Mob. Comput. Commun. Rev.*, **5**(1):3–55, January 2001. 12

[50] Claude Elwood Shannon and Warren Weaver. *The mathematical theory of communication*. University of Illinois Press, Urbana, 1949. 12

[51] T. M. Cover and Joy A Thomas. *Elements of information theory*. Wiley-Interscience, Hoboken, N.J., 2nd ed edition, 2006. 15, 16, 17

[52] A. Dembo, T.M. Cover, and J.A. Thomas. **Information theoretic inequalities**. *Information Theory, IEEE Transactions on*, **37**(6):1501–1518, 1991. 16, 17

[53] J. R Norris. *Markov chains*. Cambridge University Press, Cambridge, UK, 1st pbk. ed edition, 1998. 17

[54] G.E. Crooks. *Excursions in statistical dynamics*. PhD thesis, UNIVERSITY of CALIFORNIA, 1999. 17, 22

[55] Christopher Jarzynski. **Nonequilibrium work relations: foundations and applications**. *The European Physical Journal B - Condensed Matter and Complex Systems*, **64**(3-4):331–340, 2008. 18, 19, 23

[56] Sadi Carnot. *Reflexions sur la Puissance Motrice du Feu et sur les Machines propres à Développer cette Puissance*. 1824. 18

[57] A. Magnus. *Quaestiones Alberti de modis significandi*. Benjamins, 1977. 19

[58] G.E. Crooks. **Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems**. *Journal of Statistical Physics*, **90**(5):1481–1487, 1998. 19, 23

[59] G.E. Crooks. **Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences**. *Physical Review E*, **60**(3):2721, 1999. 20

[60] E.M.F. Curado and C. Tsallis. **Generalized statistical mechanics: connection with thermodynamics**. *Journal of Physics A: Mathematical and General*, **24**:L69, 1991. 20

[61] E.T. Jaynes. **Information theory and statistical mechanics**. *The Physical Review*, **106**(4):620–630, May 15 1957. 24

[62] E.T. Jaynes. **Information theory and statistical mechanics. II**. *Physical review*, **108**(2):171, 1957. 24

REFERENCES

[63] R. Landauer. **Irreversibility and heat generation in the computing process**. *IBM journal of research and development*, **5**(3):183–191, 1961. 26

[64] R. Landauer. **The physical nature of information**. *Physics letters A*, **217**(4-5):188–193, 1996. 26

[65] Bard Ermentrout and David H Terman. *Mathematical foundations of neuroscience*, **v. 35** of *Interdisciplinary applied mathematics*. Springer, New York, 2010. 31, 41

[66] G.E. Uhlenbeck and L.S. Ornstein. **On the theory of the Brownian motion**. *Physical Review*, **36**(5):823, 1930. 33

[67] A. Destexhe, M. Rudolph, J.M. Fellous, and T.J. Sejnowski. **Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons**. *Neuroscience*, **107**(1):13–24, 2001. 33

[68] G. La Camera, M. Giugliano, W. Senn, and S. Fusi. **The response of cortical neurons to in vivo-like input current: theory and experiment**. *Biological cybernetics*, **99**(4):279–301, 2008. 33

[69] L.C. Evans. **An introduction to stochastic differential equations**. *lecture notes, Department of Mathematics, University of California, Berkeley. http://www. math. berkeley. edu/˜ evans/SDE. course. pdf*, 2002. 34

[70] H.C. Tuckwell. **Diffusion approximations to channel noise**. *Journal of theoretical biology*, **127**(4):427–438, 1987. 35

[71] N. Fourcaud-Trocmé and N. Brunel. **Dynamics of the instantaneous firing rate in response to changes in input statistics**. *Journal of computational neuroscience*, **18**(3):311–321, 2005. 35

[72] J.J. Hopfield. **Neural networks and physical systems with emergent collective computational abilities**. *Proceedings of the national academy of sciences*, **79**(8):2554, 1982. 38

[73] H.S. Seung, T.J. Richardson, J.C. Lagarias, and J.J. Hopfield. **Minimax and Hamiltonian dynamics of excitatory-inhibitory networks**. *Advances in neural information processing systems*, **10**:329–335, 1998. 38

[74] D.R.C. Dominguez and E. Korutcheva. **Three-state neural network: From mutual information to the hamiltonian**. *Physical Review E*, **62**(2):2620, 2000. 38

[75] M.T. Wilson and D.A. Steyn-Ross. **Subthreshold dynamics of a single neuron from a Hamiltonian perspective**. *Physical Review E*, **78**(6):061908, 2008. 38

[76] L. Paninski, Y. Ahmadian, D.G. Ferreira, S. Koyama, K. Rahnama Rad, M. Vidne, J. Vogelstein, and W. Wu. **A new look at state-space models for neural data**. *Journal of computational neuroscience*, **29**(1):107–126, 2010. 38

[77] T. C Gard. *Introduction to stochastic differential equations*, **114**. M. Dekker, New York, 1988. 41

[78] D.J. Higham. **An algorithmic introduction to numerical simulation of stochastic differential equations**. *SIAM review*, pages 525–546, 2001. 41, 42

[79] W. Rümelin. **Numerical treatment of stochastic differential equations**. *SIAM Journal on Numerical Analysis*, 1982. 41

[80] Peter E Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, **23**. Springer, Berlin, corr. 3rd print edition, 1999. 42